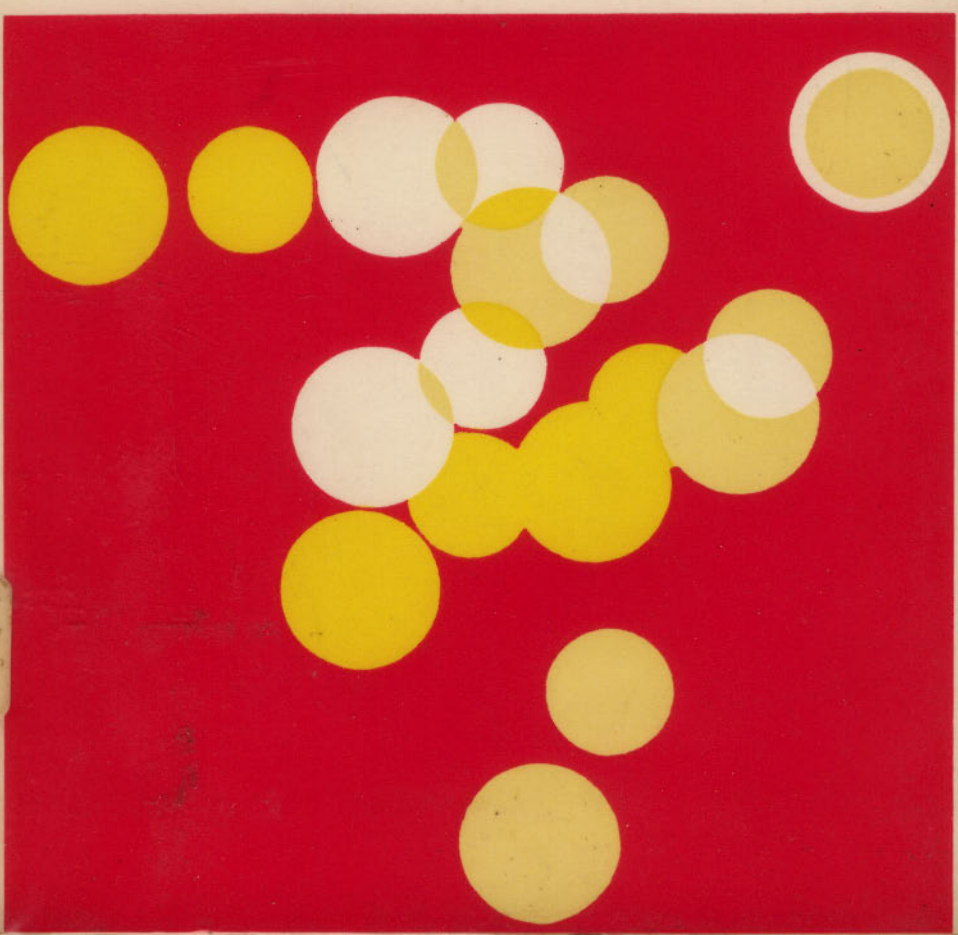
MATHEMATICAL STUDIES: 2

ALGEBRAIC STRUCTURES

A. W. BELL



A. W. BELL

ALGEBRAIC STRUCTURES

GEORGE
ALLEN
AND
UNWIN

This book is intended to form an introduction to an important branch of modern mathematics for sixth form pupils and college and university students. Its two main themes are the idea of an *axiom-system*, and the structure of the group. The idea of a set of axioms as defining an abstract mathematical structure which has a number of different concrete embodiments is central to modern mathematics; and of the variety of algebraic structures the group is probably the most important. Features of the book are the introduction of the abstract ideas through the discussion of concrete situations, and the provision of exercises which require not so much the performance of learned techniques as the investigation of further situations using the concepts which have been learned. The material is based on what has been taught for several years to mathematics students at Nottingham College of Education.

The book begins with a discussion of axiom-systems, and this is followed by a study of the familiar number systems—natural numbers, integers, rationals, real numbers—from an axiomatic point of view. From this the concepts emerge which form the basis of a study of algebraic structure—concepts like law of composition, identity and inverse element, closure. The remainder of the book is devoted to an investigation of group structure and its application. Examples of groups are discussed from a variety of fields both within and outside mathematics; finite arithmetics lead to the notion of isomorphism; permutation groups are considered, and Cayley's and Lagrange's theorems; and a chapter is devoted to applications of group theory to the symmetry of plane and solid figures and repeating patterns.

The author is Head of the Department of Mathematics at Nottingham College of Education. He is co-author of the A.T.M. Mathematics Teaching Pamphlet No. 12, *Symmetry Groups*, and a contributor to *Some Lessons in Mathematics* (edited by T. J. Fletcher).

Jacket design by Timothy Drever

Algebraic Structures
A. W. Bell
George Allen & Unwin Ltd.
Price in U.K. only: 25s. net
Paper back Edition 15s. net

MATHEMATICAL STUDIES

A Series for Teachers and Students

EDITED BY

DAVID WHEELER

School of Education, University of Leicester

No. 2

ALGEBRAIC STRUCTURES

In the same series

No. 1
TRANSFORMATION GEOMETRY
by Max Jeger

No. 3
VECTORS
by E. H. Leaton

In preparation

INTRODUCTORY TOPOLOGY by J. E. Reeve
LOGIC by W. O. Storer

Algebraic Structures

Some Aspects of Group Structure

By
A. W. BELL
College of Education, Nottingham

London
GEORGE ALLEN AND UNWIN LTD
RUSKIN HOUSE MUSEUM STREET

FIRST PUBLISHED IN 1966

PRINTED IN GREAT BRITAIN
in 10 point Times Roman type
BY C. TINLING AND CO. LTD.
LIVERPOOL, LONDON AND PRESCOT

FOREWORD

It is generally agreed that school mathematics syllabuses are in need of reform. The traditional syllabus is no longer an adequate preparation for mathematics as it is taught at a higher level; it indicates very little of the range of contemporary uses of mathematics; and it contains a high proportion of routine computation and manipulation at the expense of mathematical ideas which yield immediate enjoyment and satisfaction. A number of schools are now experimenting with new syllabuses which attempt to cure these faults.

Whether the experiments prove to be wholly successful or not, they are bringing a new element into the situation: an awareness that it is part of the job of the teacher of mathematics to inform himself about the relatively recent developments and changes in his subject. It is no longer possible to believe that developments in mathematics concern only the research mathematician and do not have any bearing on the mathematics taught in schools. This series of books is intended as a contribution to the reform of school mathematics by introducing to the reader some areas of mathematics which, broadly speaking, can be called modern, and which are beginning to have an influence on the content of school syllabuses.

The series does not put forward explicit advice about what mathematics to teach and how it should be taught. It is meant to be useful to those teachers and students in training who want to know more mathematics so that they can begin to take part in the existing experimental schemes, or modify them, or devise their own syllabus revisions, however modest. The books are elementary without being trivial: the mathematical knowledge they assume is roughly that of a traditional grammar school course, although substantial sections of all the books can be understood with less.

Now that the stability over a long period of school mathematics syllabuses seems to be coming to an end, it is to be hoped that a new orthodoxy does not succeed the old. The reform of mathematics teaching should be a continuing process, associated with a deepening study of the subject throughout every teacher's professional life. These books may help to start some teachers on that course of study.

D.W.

CONTENTS

LIST OF SYMBOLS

Page	Symbol	Meaning
17	$p \Rightarrow q$	p implies q; or if p, then q
19	$p \Leftrightarrow q$	p implies and is implied by q; or p if and only if q
17	$\sim p$	not-p
21	ε	is a member of
22	N or J^+	the set of natural numbers or positive integers
22	J	the set of integers
22	Q	the set of rational numbers
22	R	the set of real numbers
24	a^+	the successor of a
32	$a \mid b$	a divides b (exactly: meaningful only if a and b are integers)
37	^+a, ^-a	positive and negative integers
61	j	the identity function, mapping every element onto itself
61	f^{-1}	the inverse of the function f
62	$\bar{a}$	the inverse of the group-element a with respect to the composition *
70	$\mathbf{a}_m$	the residue class of integers which leave remainder a when divided by m
75	J_m	the set of residue classes modulo m
75	$J_m, +$	the group consisting of the set J_m with the composition +
76	$J_m^n, \cdot$	the group, with composition $\cdot$, of the n elements of J_m left when the zero and all elements having divisors in common with m have been excluded

LIST OF SYMBOLS

Page	Symbol	Meaning
83	(a, b, c)	the *ordered* set of elements a, b, c
65	$\{a, b, c\}$	the set consisting of the elements a, b, c $\{a, b, c\} = \{b, a, c, c\}$ but $(a, b, c) \neq (b, a, c)$
84	C_n	the cyclic group of order n
96	(abc)	the permutation $a \rightarrow b$, $b \rightarrow c$, $c \rightarrow a$
99	P_n	the group of all permutations of the set $\{1, 2, \ldots n\}$
105	A_n	the group of all *even* permutations of $\{1, 2 \ldots n\}$
115	D_n	the dihedral group of degree n

1

AXIOM SYSTEMS AND MULTIVALENT STRUCTURES

The main theme of this book is the study of group structure. This is the most fundamental of the various algebraic structures, and it leads us into some discussion of rings and fields. We do not, however, attempt to deal with vector spaces or with Boolean algebra, these being treated in other books in this series. The first part of this book is devoted to the development of the familiar number systems from an axiomatic point of view, since it is from the study of the properties of these systems that the concepts of composition, associativity, inverse element (which, with others, form the basis of the modern algebraic structures) most naturally emerge. In the second part of the book we study a number of aspects of group structure and their applications. In this introductory chapter, we consider the notions of axiom system and multivalent structure, two of the most important ideas of modern mathematics, which form themes running through the whole book.

The reader will have had some experience of mathematical proof in his study of Euclidean geometry, and perhaps also some understanding of the idea of an axiomatic mathematical system. In such a system, every theorem has to be proved using only those theorems which have been proved previously, together with the laws of logic; the first theorems of the sequence are proved from a few basic 'axioms', which are intuitively plausible and which are assumed to be true for the purpose of building up the system. Thus the theorem that the opposite angles of a cyclic quadrilateral are supplementary may be traced back, first to the theorem that the angle at the centre of a circle is twice the angle at the circumference on the same arc, then to the theorems of the exterior angle of a triangle and of the equality of the base angles of an isosceles triangle. These in turn depend on (i) the equality of corresponding and alternate angles between a transversal and a pair of parallel lines, and (ii) the congruence of two triangles which have two sides and the included angle equal. This is shown diagrammatically in Figure 1. Statements (i) and (ii) may be taken as axioms—they appear plausible and we do not attempt to prove them.

Exercise

Choose three geometrical theorems and trace back their proofs: try to show that they may all be proved starting from the axioms (i) and (ii) above.

The result of this exercise is that a theorem such as the one quoted about the cyclic quadrilateral carries the same degree of conviction as the axioms from which we started: if, in any situation, we are convinced that the axioms apply, we must be equally convinced that the theorem will be true. However, two points remain to be discussed. The first of these is the question of definitions. It is easy enough to define 'circle' or 'triangle' or 'congruent' in terms of more basic words like 'point', 'line', 'distance', but it is difficult, if not impossible, to define these basic concepts in terms of anything more fundamental. Euclid said, 'A point is that which has no size', and 'A line is length without breadth', but these statements do not actually tell us what these things are, nor can they be used as the basis of proofs of their properties. But although we cannot satisfactorily define 'point' or 'line', we can make certain statements about them which we may take as axioms and which would be usable in proving theorems: for example, 'Any two points lie on a unique straight line', and 'Two straight lines meet in not more than one point'. 'Point' and 'line' remain as *undefined terms.* The standard pattern for a mathematical system then becomes—

undefined terms;
axioms stating the relationships among these terms;
definitions of further terms;
theorems stating relationships deduced from the above.

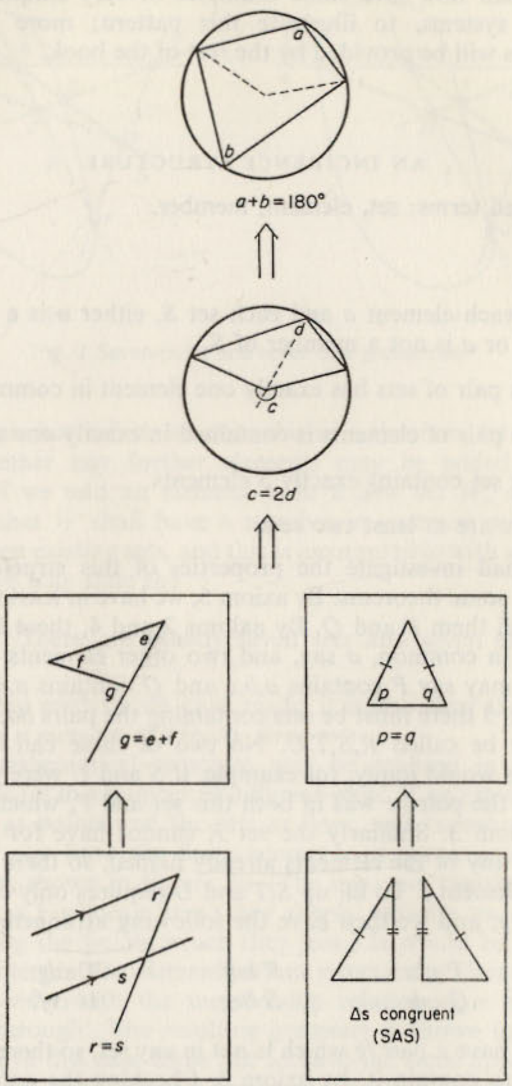


Fig. 1 Analysis of the proof of the cyclic quadrilateral theorem.

We shall now give some examples of very simple mathematical systems, to illustrate this pattern; more complex examples will be provided by the rest of the book.

AN INCIDENCE STRUCTURE

Undefined terms: set, element, member.

Axioms

1. For each element *a* and each set *S*, either *a* is a member of *S* or *a* is not a member of *S*.
2. Each pair of sets has exactly one element in common.
3. Each pair of elements is contained in exactly one set.
4. Each set contains exactly 3 elements.
5. There are at least two sets.

We shall investigate the properties of this structure, and arrive at some theorems. By axiom 5, we have at least two sets: let us call them *P* and *Q*. By axioms 2 and 4, these have one element in common, *a* say, and two other elements each, so that we may say *P* contains *a,b,c* and *Q* contains *a,d,e*. Now by axiom 3 there must be sets containing the pairs *bd,be,cd,ce;* let these be called *R,S,T,U*. No two of these can coincide, since this would imply, for example, if *S* and *U* were the same set, that the pair *bc* was in both this set and *P*, which contradicts axiom 3. Similarly the set *R* cannot have for its third member any of the elements already named, so there must be a sixth element *f*. To fill up *S,T* and *U* requires only one more element *g*, and we then have the following arrangement:

P abc	*R bdf*	*T cdg*
Q ade	*S beg*	*U cef.*

We now have a pair *fg* which is not in any set, so there must be a set *V* to contain it, by axiom 3. Checking the pairs in the scheme above shows that the third element of this set must be *a*. This system of seven elements and seven sets satisfies all

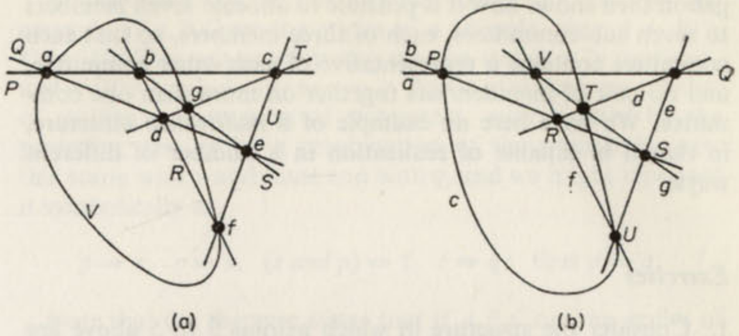


Fig. 2 Seven-point and seven-line geometries.

the axioms, but before we state this as a theorem, let us consider whether any further elements may be added to the system. If we add an element *h*, in a new set *W*, axiom 2 requires that *W* shall have a member in common with each of the seven existing sets, and this is incompatible with axiom 4. Thus we may state the

Theorem There are exactly seven sets and seven elements.

From the analysis we have made, it is also clear that *every element is a member of exactly three sets.*

This mathematical structure may be realized in a more concrete form in a number of different ways. If we interpret the elements as points and the sets as lines, and 'member of' as 'lying on', we have the finite geometry of seven points and seven lines shown in Figure 2a. (The lines can clearly not be regarded as Euclidean lines; they simply indicate an association among the points which they join.) It would be equally valid to interpret the elements of our structure as *lines* and the sets as *points*, with the membership relation now meaning 'passing through'. The resulting geometry is shown in Figure 2b, which in this case turns out to have the same appearance as Figure 2a. (This is a *self-dual* configuration.) Another interpretation would be to regard the elements as members of a main committee, and the sets as sub-committees. Our investigation then shows how it is possible to allocate seven members to seven sub-committees, each of three members, so that each committee contains a representative of each other committee, and no pair of members sits together on more than one committee. We have here an example of a *multivalent* structure, in that it is capable of realization in a number of different ways.

Exercises

1. Consider the structure in which axioms 3 to 5 above are replaced by the following:

 3. Each element is a member of exactly two sets.
 4. There are four sets.

Show that there are exactly six elements; and, defining a pair of *disjoint* elements as a pair not contained in any set, show that each element has exactly one element with which it is disjoint.

Interpret this structure, and the theorems, in terms of points and lines in two different ways, and in terms of committee members. Are the resulting geometrical configurations self-dual, in this case?

2. Vary axiom 4 in the above exercise, and investigate the resulting systems.

3. (Harder) Look up Desargues' Theorem and Pappus' Theorem in a textbook of projective geometry, and try to design a set of axioms to give each of these configurations.

LOGIC

We now take up the second of the two points which arose from the discussion of the deductive development of Euclidean geometry. This is the question of the rules of logic by which one statement can be said to 'follow from' a previous one. We shall not discuss these exhaustively, but will comment on a few points of importance and introduce a few symbols which will be of use later in the book.

All theorems can be put into the form 'If p, then q'; for example, *If* A,B,C are the angles of a triangle, *then* $A + B + C = 180°$.This relationship may also be stated as 'p *implies* q', and written symbolically as $p \Rightarrow q$. Proving the theorem consists of making a succession of statements, each implied by the preceding one or by a combination of the preceding ones; this starts with p and must end with q, and we might represent it symbolically as

$$p \Rightarrow r, \quad r \Rightarrow s, \quad (s \text{ and } p) \Rightarrow t, \quad t \Rightarrow q; \quad \text{thus } p \Rightarrow q.$$

Note that the theorem states that *if* A,B,C are the angles of a triangle, *then* $A + B + C = 180°$: it says nothing about what may be the case if A,B,C are *not* the angles of a triangle. The ordinary usage of these connectives is not always clear. If, for example, I say, 'If it rains tomorrow I shall not go to the match', there is a possible underlying implication that if it does not rain I shall go. Further confusion is caused by the fact that many mathematical theorems are actually true both ways round; for example, if a quadrilateral is a parallelogram then its diagonals bisect one another, and it is also true that if a quadrilateral is not a parallelogram then its diagonals do not bisect each other. We must agree to use the terms in the strict sense that 'If p, then q' is a statement about what follows *if p is true*, and that it says nothing at all about what may be the case if p is not true.

The statement, 'If A,B,C are not the angles of a triangle, then $A + B + C \neq 180°$,' is called the *opposite* of the original statement: symbolically, the opposite of $p \Rightarrow q$ is not-p $\Rightarrow$ not-q, or using the accepted symbol $\sim p$ for not-p, the opposite is $\sim p \Rightarrow \sim q$. We see that the opposite does not follow from the original theorem. There are two further statements which are logically related to an implication; the *converse*, $q \Rightarrow p$, and the *contrapositive*, $\sim q \Rightarrow \sim p$. Figure 3 displays the four related statements for the case under discussion. It is clear that the contrapositive is true, and the converse false; in this case, moreover, the contrapositive is equivalent to the original theorem, and the converse and the opposite are equivalent to each other. Reflection on the meaning of the statements should make this obvious; it can also be demon-

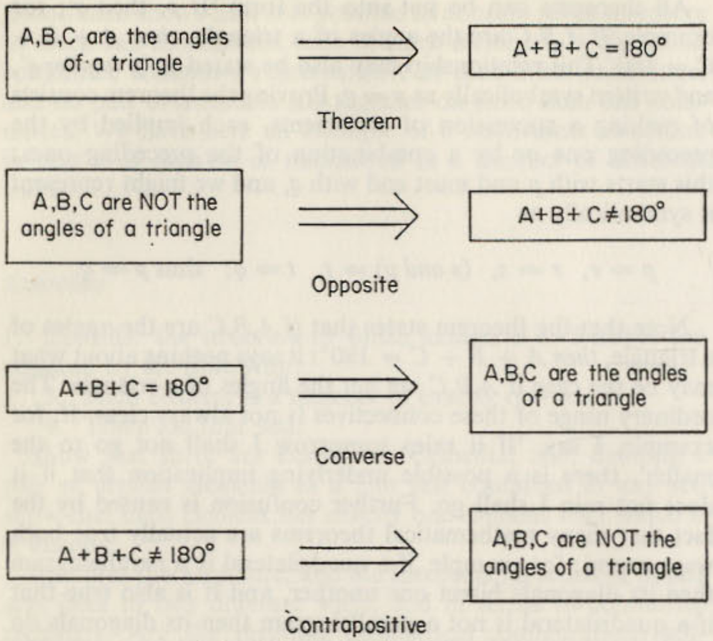

A,B,C are the angles of a triangle
A+B+C = 180°
Theorem
A,B,C are NOT the angles of a triangle
A+B+C ≠ 180°
Opposite
A+B+C = 180°
A,B,C are the angles of a triangle
Converse
A+B+C ≠ 180°
A,B,C are NOT the angles of a triangle
Contrapositive

Fig. 3 Four related statements.

strated symbolically. To do this, it is convenient to use the fact that

$$p \Rightarrow q$$

is equivalent to the statement that

$$(p \text{ and } \sim q) \text{ is false,}$$

for p being true, and q not, is the one case excluded by the statement $p \Rightarrow q$. Then

$$\sim q \Rightarrow \sim p$$

is equivalent to

$$(\sim q \text{ and } \sim(\sim p)\,) \text{ is false,}$$

which is equivalent to

$$(\sim q \text{ and } p) \text{ is false,}$$

which is equivalent to

$$p \Rightarrow q;$$

and similarly, we have $q \Rightarrow p$ equivalent to

$$(q \text{ and } \sim p) \text{ is false,}$$

and $\sim p \Rightarrow \sim q$ equivalent to

$$(\sim p \text{ and } \sim(\sim q)\,) \text{ is false,}$$

so that $q \Rightarrow p$ is equivalent to $\sim p \Rightarrow \sim q$.

We shall use one other example to illustrate these ideas—the theorem of Pythagoras.

Theorem: $p \Rightarrow q$: If $\triangle ABC$ is right-angled at A, then $BC^2 = AB^2 + AC^2$.

Opposite: $\sim p \Rightarrow \sim q$: If $\triangle ABC$ is not right-angled at A, then $BC^2 \neq AB^2 + AC^2$.

Converse: $q \Rightarrow p$: If $BC^2 = AB^2 + AC^2$, then $\triangle ABC$ is right-angled at A.

Contrapositive: $\sim q \Rightarrow \sim p$: If $BC^2 \neq AB^2 + AC^2$, then $\triangle ABC$ is not right-angled at A.

In this case, both the theorem and the converse can be proved true; we have $p \Rightarrow q$ and $q \Rightarrow p$, which is written $p \Leftrightarrow q$, and we say that *p implies and is implied by q*, or p is equivalent to q. Other common forms of words are 'p if and only if q' (sometimes abbreviated to 'p iff q'), and 'p is a necessary and sufficient condition for q'.

Exercise

Consider the statement 'If ABC is a triangle, acute-angled at A, $BC^2 \neq AB^2 + AC^2$.' Write its converse, opposite and contrapositive, and consider their truth.

Reductio ad absurdum

This type of proof, which is quite commonly used, consists of proving a statement by showing that the assumption that it is false leads to a contradiction. For example, to prove that if points C, D, on the same side of AB, are such that $\angle ACB = \angle ADB$, then the four points A, B, C, D are concyclic, we may start by assuming that the circle through A,B,C does *not* contain D, and show that this leads to a contradiction. In terms of the symbols used above, we are proving that $p \Rightarrow q$ by showing directly that (p and $\sim q$) is impossible. Another example of this type of proof may be found on p. 52 of this book, where we prove that if $x^2 = 2$, then x is not a rational number by showing that, $x^2 = 2$ and x is rational, is impossible.

Exercise

Prove the converse of Pythagoras' Theorem by a *reductio ad absurdum* method, assuming the truth of Pythagoras' Theorem itself.

Quantifiers

Mathematical theorems are often generalizations which are true for all members of some set; this is often below the surface even when it is not explicitly stated. The two theorems discussed above, for example, are true *for all triangles*. This point is sometimes important, as we shall see in this example:

$$\text{If } x^2 = y^2 \text{ then } x = y.$$

Whether this statement is true or not cannot be discussed until we know what the x and y stand for. The meaning might be one of the following:

(i) For all x,y in the set of natural numbers,

$$x^2 = y^2 \Rightarrow x = y.$$

(ii) For all x,y in the set of integers (positive and negative),

$$x^2 = y^2 \Rightarrow x = y.$$

We can now say that the first of these theorems is true, and the second false.

Note that to disprove the second theorem it is only necessary to state one pair x,y for which it is not true, for instance, 3, -3 (this is called disproof by counter-example). In other words, the negation of 'for all x,y in the set N, $p \Rightarrow q$', is 'for *some* x,y in the set N, $p \not\Rightarrow q$' (p does not imply q)—and the 'some' may be only one; or '*there exists* a pair x,y in the set N. . .' The *quantifiers* 'for all' and 'for some' are denoted by the symbols $\forall$ and $\exists$, and if we use also the symbols N for the set of natural numbers (positive integers), J for the whole set of integers, positive, negative and zero, and ε for 'is a member of', the statements under discussion would read

$$\forall x,y \ \varepsilon \ N, x^2 = y^2 \Rightarrow x = y,$$

which is true, and

$$\forall x,y \ \varepsilon \ J, x^2 = y^2 \Rightarrow x = y,$$

which is false, since

$$\exists \ x,y \ \varepsilon \ J, x^2 = y^2 \text{ and } x \neq y.$$

Exercises

Discuss the following statements, using all the relevant ideas from this section on logic.

1. If a^2 is even, then a is even.
2. $n\,(n + 1)\,(2n + 1)$ is divisible by 6.
3. The sides of a square are all equal.
4. $n^2 - 1 = (n + 1)\,(n - 1)$.
5. $n^2 - 1 = 3$.
6. There is an infinite number of prime numbers.

REFERENCES

ALLENDOERFER, C. B. and C. O. OAKLEY, *Principles of Mathematics*, Chapter 1, McGraw-Hill, New York, 2nd edn, 1963.

STABLER, E. R., *An Introduction to Mathematical Thought*, Addison-Wesley, New York, 1953.

SMITHIES, F., 'What is modern mathematics', *Mathematical Gazette*, December 1963.

2

NATURAL NUMBERS

We have stated in Chapter 1 that one of the key ideas of modern mathematics is the multivalent structure, and the chief such structures which we shall discuss are function, law of composition, group, ring and field. These are, however, generalizations which have arisen historically out of the study of more familiar univalent structures, such as the number systems and the cartesian (coordinate) plane; the origins of the incidence structures of the previous chapter, in the geometry of points and lines, were fairly obvious, and the same will be seen to apply to the other structures we shall consider.

We shall therefore begin our study of algebraic structures with the number systems in common use. This will also have the advantages of giving the reader experience of the axiomatic method applied to familiar objects, and of helping him to clarify his understanding of number systems at several points which cannot be fully dealt with at a more elementary level.

First we must distinguish between the different types of numbers which we use. The counting numbers 1,2,3, . . . are called *natural numbers* and we shall denote this set by the symbol N. The set 0, ± 1, ± 2, . . . is the set of *signed integers*, or simply the *integers*, and will be denoted by J. The set of numbers of the form p/q, where p and q are integers, with $q \neq 0$, will be called the set of *rational numbers*, and denoted by Q; the subset of positive rationals will be denoted by Q^+. The set containing all these, and also all numbers which can be expressed as infinite decimals, such as $\sqrt{2}, \pi$, $\sqrt{(\sqrt{3} + 1)}$ and so on, is the set of *real numbers*, for which the symbol R will be used. Other types of number will be defined at appropriate places in the text.

NATURAL NUMBERS

In our attempt to define the different types of number as precisely as possible, the first step will prove to be one of the most difficult to take. Once the natural numbers are defined and their properties proved, it is a comparatively easy matter to define the integers and the rationals in terms of them, though the real numbers present somewhat greater difficulty. We shall start by examining the properties of the natural numbers, and then state a set of axioms by which they may be defined.

In ordinary usage, the word *number* may refer either to one of the sequence of symbols 1,2,3, . . . or to a property of a set of objects, for example the property of having a one-one correspondence between its objects and the symbols 1,2 and 3, of the sequence above. When we wish to make this distinction, we shall call a symbol like 3 or 26 a *number-symbol* or *numeral*, and for the other meaning we shall use the phrase the *number of a set*. A child's first step towards counting consists of learning the set of number-words, and their order. (Their order is, in fact, the only significant thing about these; they are simply a set of words possessing an order which is universally agreed.) The next step consists of learning to make a one-one correspondence between the words and the objects which are to be counted or numbered. Counted *or* numbered—for there are two distinct uses to which this sequence of number-symbols may be put. They may be used to *count* the number of objects in a set —this is called a *cardinal* use of the number system—or to *identify* a particular member of a sequence, such as a sequence of houses in a street, or of kings of the same name—this is called an *ordinal* use of the numbers.

We shall now state a set of axioms for a natural number sequence. These were first given by the Italian mathematician Peano, and it will be seen that they are mainly a precise statement of the property of order which we have already seen to be the essential property of the natural number sequence. Care has to be taken to ensure that what we allow is a single chain-like sequence, without any branching.

PEANO'S AXIOMS FOR A NATURAL NUMBER SEQUENCE

A *natural number sequence* is a set N of elements together with a successor function $(^+)$ satisfying the following axioms:

N1. For each element a of N, there is a unique element a^+ in N, called the successor of a.

N2. No two elements a,b have the same successor.

N3. N contains a unique element 1 which is not the successor of any element of N.

N4. Every set S of elements of N such that
(a) 1 is in S, and
(b) if a is in S, then its successor a^+ is in S,
consists of the whole set N.

Any member of a natural number sequence is called a natural number. By this definition, the set 1,2,3, . . . 9,10, . . . of numerals in base ten is a natural number system: so is the set 1,2,10,11,12,20, . . . of numerals in base three. We now go on to define the number of a set, but first we introduce a useful term.

Definition A segment $\langle 1\ \mathrm{n} \rangle$ of a natural number sequence N is a subset S of N such that
(a) S contains 1, and
(b) if $x \varepsilon$ S, $x^+ \varepsilon$ S; unless $x = n$, when $x^+ \notin S$.

The number of a set

A set is said to have number n (or to *contain n elements*) if its elements can be placed in one-one correspondence with the members of the segment $\langle 1\ \mathrm{n} \rangle$ of a natural number sequence.

A set whose elements may be put into one-one correspondence with the *whole* of a natural number sequence is said to have number ω.

We have mentioned that the numerical sequences in different bases may all be described as 'natural number sequences', and we therefore need to provide some definition of equality for two numbers, to make the necessary correspondence between two different sequences.

Definition The numbers m,n of two natural number sequences are said to be *equal* if the segments $\langle 1\ m \rangle$ and $\langle 1\ n \rangle$ may be put in one-one correspondence with each other.

PROPERTIES OF NATURAL NUMBERS

We have given a set of axioms for 'natural numbérs' and have shown that the familiar set of numerals 1,2,3,4, . . . satisfies these axioms. We have also shown how a suitable member of a natural number sequence may be assigned to a set, to be called the number of the set. What we have not so far considered is whether all the familiar properties of natural numbers, as we know them, may be proved from the axioms *N1* to *N4*. This raises the question, what are these 'familiar properties'? To answer this we shall have to examine the way in which we habitually use numbers, and try to note what the laws are which we use unconsciously most of the time.†

Consider the following calculations:

$$\begin{array}{r} 28\\ +34\\ +42\\ \hline 104\\ \hline \end{array} \qquad \begin{array}{r} 682\\ -324\\ \hline 358\\ \hline \end{array} \qquad \begin{array}{r} 35\\ \times 23\\ \hline 105\\ 70\\ \hline 805\\ \hline \end{array} \qquad \begin{array}{r} 12 \text{ rem } 9\\ 23\overline{)285}\\ 23\\ \hline 55\\ 46\\ \hline 9\\ \hline \end{array}$$

First, we note that there are four ways of combining numbers, of which two, addition and multiplication, are in some senses primary processes and the other two, subtraction and division, are derived from them: any two numbers may be added or multiplied, but 5 cannot be subtracted from 3 nor divided by 3, without inventing new types of number. This raises the second point, which is that $a + b$ and $b + a$ are equal, but $a - b$ and

† The question of whether the Peano axioms specify uniquely the set of natural numbers is a difficult one; it is generally agreed that they do not. Waismann (*Introduction to Mathematical Thinking*) says that 'Skolem has proved . . . that no one will ever succeed in characterising the set of integers with a finite group of axioms.' (p. 148).

$b - a$ are not, and similarly $ab = ba$ but $a \div b \neq b \div a$. Next, although addition is in the first place a way of combining *two* numbers, three or more numbers may be added, and the result is the same whichever pair is combined first: $a + (b + c) = (a + b) + c$: in the addition shown above, we may say $(8 + 4) + 2 = 12 + 2 = 14$, or $8 + (4 + 2) = 8 + 6 = 14$. The same law applies to multiplication, but not, of course, to subtraction or division. Then, in the multiplication, we use the fact that 35 multiplied by 23 may be expressed as $(35 \times 20) + (35 \times 3)$: in general, $a(b + c) = ab + ac$. The basis of this law is the fact that multiplication is repeated addition: 3×35 means $35 + 35 + 35$. In the division we use a more complicated fact, that if b is less than a (less than—another property to define), then b may be divided either into a, or into some number less than a, leaving a remainder, which is less than b: in symbols, if $b < a$, there are numbers q and r, with $r < b$, such that $(a - r) \div b = q$, that is $a = bq + r$. Finally, we have introduced an extra number into the system, the zero, whose properties are that, for any number x, $x + 0 = x$ and $x.0 = 0$; and there is a further pair of laws which we use, and which do not follow logically from the others—the cancellation laws, which state that if $a + c = b + c$, then $a = b$, and if $ac = bc$, and $c \neq 0$, then $a = b$. To summarize, we have to make definitions and prove properties as follows, in terms of the axioms we have stated.

Definitions

$+, -, \times, \div$, (and also powers and roots, for completeness), less than, 0.

Properties

Commutative laws:

$$a + b = b + a, \quad ab = ba.$$

Associative laws:

$$a + (b + c) = (a + b) + c, \quad a(bc) = (ab)c.$$

Distributive law:

$$a(b + c) = ab + ac.$$

Cancellation laws:

$$a + c = b + c \Rightarrow a = b, \quad ac = bc \text{ and } c \neq 0 \Rightarrow a = b.$$

Properties of zero:

$$a + 0 = a, \quad a.0 = 0.$$

Division theorem: if $b < a$, there are numbers q,r with $r < b$, such that

$$a = bq + r.$$

Exercise

Write out in full the steps of the four calculations, noting at each step which of the properties you use.

SUM OF TWO NATURAL NUMBERS

The usual way of finding the sum of two numbers, such as 4 and 3, is to count up to the 4, and then three places more in the same sequence—we make a one-one correspondence between the 5,6,7 of the first sequence and the 1,2,3 of a second sequence, as in Figure 4a.

$$\begin{array}{l} 1,2,3,4,5,6,7,\ldots \\ \qquad\quad 1,2,3,\ldots \end{array} \qquad\qquad \begin{array}{lll} 1,\ldots a, a^+,\ldots & a + b, & (a + b)^+ \\ \qquad 1,\ldots & \quad b & \quad b^+ \end{array}$$

(a) (b)

Fig. 4

Figure 4b shows the situation in terms of general numbers a, b. Here a one-one correspondence is made between the segment $\langle 1\, b \rangle$ and a similar 'segment' starting with the successor of a. The end point of this 'segment' is the number we call $a + b$, and we might take this as a definition of $a + b$. However, it is easier to use a definition which avoids the notion of segment, and is based directly on the terms contained in the axioms. Such a definition has to be a recursive one, that is it does not define $a + b$ directly, but defines $a + 1$, and then tells us how to find $a + b^+$ if we know $a + b$. The reason for this is that these are the only two relations we can derive from the addition process as shown in Figure 4b: it is clear there that $a + 1 = a^+$, and that $a + b^+ = (a + b)^+$.

Definition The *sum* $a + b$ of two natural numbers is defined by the following:

$$a + 1 = a^+, \tag{1}$$

$$a + b^+ = (a + b)^+. \tag{2}$$

Mathematical Induction

Having defined addition in this way, we shall find that to prove the properties set out above we make extensive use of the method of proof by mathematical induction, which means simply that we shall be making use of the axiom *N4*. This is not really surprising: what our axioms define is essentially an unending chain of 'successors', and it is therefore natural that the methods of proof used should make use of the successor idea. The method of proof by induction has the following pattern: a statement $S(n)$, containing n, is to be proved true when n is any natural number. It is first shown that the statement is true when $n = 1$; then it is shown that *if* there is any one value k of n, for which $S(n)$ is true, *then* it is also true for the successor of k, k^+. We then appeal to the induction axiom *N4*, and consider the subset T of N which contains all the values of n for which $S(n)$ is true. We know this subset T contains 1, and that *if* it contains any element k, *then* it contains k^+; we can therefore deduce from axiom *N4* that T consists of the whole set N of natural numbers, that is to say $S(n)$ is true when n is any natural number. Thus, to summarize,

$$\left.\begin{array}{ll} & S(1) \text{ is true} \\ \text{and} & \\ & S(k) \Rightarrow S(k^+) \end{array}\right\} \Rightarrow S(n) \text{ true for all } n \varepsilon N.$$

Exercises

Prove by induction (a) $1 + 2 + 3 + \ldots + n = \frac{1}{2}n(n + 1)$, (b) the amount at compound interest $r\%$ from a principal P for n years is $P(1 + r/100)^n$.

The Associative Law for Addition

For all natural numbers a, b, c,

$$a + (b + c) = (a + b) + c. \tag{3}$$

Proof We prove this by induction on c, that is to say we consider it as a statement $S(c)$ which has to be proved true when c is any natural number.

We know that $S(1)$ is true; for this states that

$$a + (b + 1) = (a + b) + 1,$$

which, by (1) of the definition of addition, is

$$a + b^+ = (a + b)^+,$$

and this is true, by (2) of the definition.

Now we show that $S(k) \Rightarrow S(k + 1)$; that is to say,

$$\begin{array}{lll} \text{if} & a + (b + k) = (a + b) + k, & (4) \\ \text{then} & a + (b + k^+) = (a + b) + k^+ & (5). \end{array}$$

$$\begin{array}{llll} \text{We have} & a + (b + k^+) & = a + (b + k)^+ & \text{by (2)} \\ & & = [a + (b + k)]^+ & \text{by (2)} \\ & & = [(a + b) + k]^+ & \text{by (4)} \\ & & = (a + b) + k^+ & \text{by (2).} \end{array}$$

Thus $(4) \Rightarrow (5)$.

Thus the conditions of the induction principle are fulfilled, and we may deduce that $S(c)$ is true for all $c \varepsilon N$, that is to say,

$$a + (b + c) = (a + b) + c \text{ for all } a,b,c \varepsilon N.$$

The Commutative Law for Addition

For all $a,b \varepsilon N$,

$$a + b = b + a. \tag{6}$$

The proof of this is left as an exercise for the reader. It involves two applications of the induction principle, one to prove that $a + 1 = 1 + a$ for all $a \varepsilon N$, and a second to deduce from this the result for all $a,b \varepsilon N$.

Cancellation Laws for Addition

For all $a,b,c \varepsilon N$,

$$a + c = b + c \Rightarrow a = b.$$

This too may be proved by induction on c.

We have $\quad a + 1 = b + 1 \Rightarrow a = b, \quad$ (this is $S(1)$)

for by (1) above this is $a^+ = b^+ \Rightarrow a = b$, which is true by the axiom *N2*.

Also, if there is a k such that

$$a + k = b + k \Rightarrow a = b, \qquad (\text{this is } S(k))$$

$$\begin{aligned} \text{we have } a + k^+ = b + k^+ &\Rightarrow (a + k)^+ = (b + k)^+ && \text{by (2)} \\ &\Rightarrow a + k = b + k && \text{by } N2 \\ &\Rightarrow a = b && \text{by } S(k). \end{aligned}$$

Thus $\mathrm{S}(k) \Rightarrow S(k^+)$ and, by the induction principle,

$$a + c = b + c \Rightarrow a = b \quad \text{for all } a,b,c \varepsilon N.$$

SUBTRACTION AND GREATER AND LESS

These are defined in terms of addition, as follows:

Definition If, for any two natural numbers a and b, there is a natural number c such that $b + c = a$, we call c the *difference* between a and b and write $c = a - b$; we also say that a is *greater than* b, $a > b$, and b is *less than* a, $b < a$.

The most important properties of the relation $<$ are contained in the following theorems:

Transitivity of $<$ If $a < b$ and $b < c$, then $a < c$.

Proof $\quad a < b \Rightarrow \exists x$† such that $a + x = b$,
$\quad b < c \Rightarrow \exists y$ such that $b + y = c$.

† 'there exists an x'.

$$\begin{aligned} \text{Then} \qquad c &= b + y \\ &= (a + x) + y \\ &= a + (x + y). \end{aligned}$$

Hence $\quad a < c.$

Law of trichotomy For all $a,b \varepsilon N$,

either $\quad a = b, \quad$ or $\quad a < b, \quad$ or $\quad b < a.$

Like most other theorems in this chapter, this may be proved by induction, but in this case the details are tedious; we shall omit this proof.

Associativity Relations for Subtraction

The associative law does not hold for subtraction, but we frequently need to use expressions in which three or more terms are combined by additions or subtractions. The fundamental relations needed are the following. They may all be proved in a few lines from the definition of subtraction.

$$\begin{aligned} a - (b - c) &= (a - b) + c, \\ a - (b + c) &= (a - b) - c, \\ a + (b - c) &= (a + b) - c. \end{aligned}$$

In practice such expressions are usually converted by using signed numbers (see Theorem 2 on p. 40 below), thus making all the operations addition, so that the associative law holds.

MULTIPLICATION

Figure 5 shows diagrammatically how multiplication, as repeated addition, fits in with the idea of a number sequence. We have a set of consecutive groups, of length equal to b, the groups being in one-one correspondence with the segment $\langle 1\ a \rangle$; the end-point of the whole sequence is what we call ab.

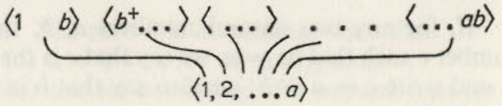


Fig. 5

This diagram gives us the essentials of a recursive definition for ab, similar to that given for $a + b$:

Definition For all $a,b \varepsilon N$,

$$1.b = b \tag{7}$$

and

$$a^+.b = ab + b. \tag{8}$$

This time we shall prove the distributive law from the definition.

Distributive Law For all $a,b,c \varepsilon N$,

$$a(b + c) = ab + ac. \tag{9}$$

Proof This is by induction on a.

The law is true for all b, c when $a = 1$, for

$$\begin{aligned} 1(b + c) &= b + c && \text{by (7)}\\ &= 1.b + 1.c && \text{by (7).}\end{aligned}$$

Also if, for some k,

$$k(b + c) = kb + kc, \tag{10}$$

we have

$$\begin{aligned} k^+(b + c) &= k(b + c) + (b + c) && \text{by (8)}\\ &= (kb + kc) + (b + c) && \text{by (10)}\\ &= (kb + b) + (kc + c) && \text{by (3) and (6)}\\ &= k^+.b + k^+.c && \text{by (8).}\end{aligned}$$

Thus the two conditions of the induction principle are satisfied and the result follows.

Exercises

Prove the associative, commutative and cancellation laws for multiplication.

DEFINITION OF DIVISION

This follows the same pattern as the definition of subtraction.

Definition If, for any two natural numbers a, b, there is a natural number c such that $bc = a$, we say that c is the *quotient* of a by b, and write $c = a \div b$; we also say that b is a *divisor* of a, and write $b \mid a$.

Like the relation 'less than', the 'divides' relation, $b \mid a$, is transitive: that is, $(c \mid b \text{ and } b \mid a) \Rightarrow c \mid a$. (The proof is left to the reader.) But these relations differ in that, while for any two natural numbers a, b, either $a < b$ or $b < a$ or $a = b$, in the case of the divisor relation we may have numbers a, b for which neither $b \mid a$ nor $a \mid b$ is true. It is this fact which is at the basis of most of the subject known as Number Theory, the study of prime numbers and divisibility. It also makes relevant the Division Theorem, which states that if b does not divide a, it does divide some number between a and $a - b$. This is the theorem on which the standard division calculation is based.

Division Theorem If $a,b \varepsilon N$ and $b < a$, then either $b \mid a$ or there exist $q,r \varepsilon N$, with $r < b$, such that $a = bq + r$.

Proof (by induction on a). The theorem is true when $a = 2$, for then $b = 1$ and $b \mid a$. Suppose, for some k, the theorem is true, that is, for all $b < k$ either

$$b \mid k \quad \text{or} \quad k = qb + r, \quad \text{with } r < b.$$

Then, if $b \mid k$, we have $k^+ = qb + 1$, which is of the required form provided $b > 1$, and if $b = 1$, we have $b \mid k^+$ in any case.

On the other hand, if $k = bq + r$, we have $k^+ = qb + (r + 1)$, which is of the required form unless $r + 1 = b$, in which case $b \mid k^+$.

Thus, in all cases, if the theorem is true for some k, it is true for k^+; hence, by the induction principle, it is true for all $a \varepsilon N$, except possibly $a = 1$.

DEFINITION OF ZERO

We noted above that zero possesses an additive and a multiplicative property. We use the first of these to define it, and prove the second.

Definition The zero 0 is an element such that

(a) $a + 0 = a$ for all $a \varepsilon N$, and (11)

(b) the set consisting of a natural number sequence and 0 has all the previously proved properties.

A natural number sequence, with zero, will be designated by the symbol N^0. It may be readily proved that there cannot be two different zero elements in a system.

We now proceed to prove that, for all $a \varepsilon N$, $0.a = 0$.

Proof For all $a \varepsilon N$,

$$\begin{aligned} a.a + a.0 &= a\,(a+0) && \text{by (9)} \\ &= a.a && \text{by (11)} \\ &= a.a + 0 && \text{by (11).} \end{aligned}$$

Thus $\qquad a.0 = 0$

by the cancellation law for addition.

POWERS AND ROOTS

Just as repeated addition is sufficiently important to be described as a distinct way of combining two numbers, so we come at a later stage to give a name to repeated multiplication; this operation of raising to a power may be defined in similar terms to addition and multiplication, as follows:

Definition The bth power of a, a^b, is defined by

$$\text{(a) } a^1 = a, \qquad (12)$$

$$\text{(b) } a^{b^+} = a^b.a. \qquad (13)$$

The commutative and associative laws do not apply to this operation, but there are two important 'laws of indices' which may be proved from the definitions by methods similar to those used in the rest of this chapter. We may also complete the definition by showing that $a^0 = 1$ for all $a \varepsilon N$.

Laws of Indices For all $a,b,c \ \varepsilon \ N$,

$$a^{b+c} = a^b.\,a^c \qquad (14)$$

and $$a^{bc} = (a^b)^c. \qquad (15)$$

The proofs of these are left to the reader.

Proof of $a^0 = 1$ for all $a \varepsilon N$.

We have

$$\begin{aligned} 1.a^b &= a^b && \text{by (7)} \\ &= a^{b+0} && \text{by (11)} \\ &= a^b.\,a^0 && \text{by (14)} \\ &= a^0.\,a^b \end{aligned}$$

by the commutative law for multiplication.

Thus $\qquad 1 = a^0$

by the cancellation law for multiplication.

Definition of root If, for two natural numbers a,b, there is a natural number c such that $c^b = a$, then c is called the bth root of a, written $c = \sqrt[b]{a}$.

3

EXTENSIONS OF THE NUMBER SYSTEM—INTEGERS, RATIONALS AND REAL NUMBERS

The need to extend the system of natural numbers to form the integers arises typically in two ways—from applications, and from within mathematics itself. The scale shown in Figure 6, with numbers marked in both directions from a central zero, is used for measuring temperatures, and a similar number system is required for recording bank balances, or the heights of people above or below a standard height. Within mathematics, manipulation is eased if two numbers can always be subtracted one from the other; but signed numbers also produce the greater advantage that any subtraction can be replaced by an addition, and vice versa. The integers consist essentially of a double set of natural numbers, with some means of distinguishing the two component sets, and a suitable definition of the relation between them. One may use + and − signs, or leave one set unsigned, and use a minus sign to distinguish the other, but it would be quite valid to call one set R1, R2, R3, . . . and the other L1, L2, L3, . . . or to use any other distinguishing marks. Thus, when we have formulated and stated a set of axioms for these numbers, we shall need to find and prove theorems governing the ways in which the signs affect the combination of these numbers; in other aspects of their manipulation they will follow the same laws as natural numbers do.

The ways in which signed numbers combine may be observed by considering the scale of Figure 6(i). Like natural numbers, the integers may be used cardinally or ordinally. When used for recording a temperature or a bank balance, they simply indicate a point on the scale: this is an ordinal use, and there is no question of adding two such numbers—one does not

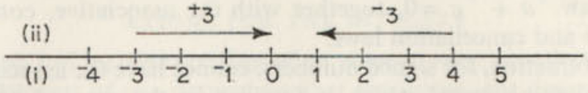
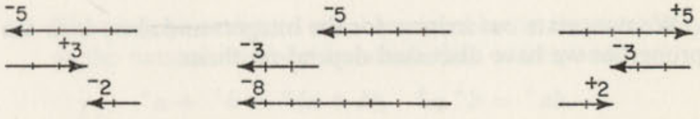


Fig. 6 Signed numbers used to describe (i) positions on a scale and (ii) changes of position.

normally add two temperatures or two points of a scale. But if it is desired to measure *changes* in temperature, or deposits in the bank, again a set of signed numbers is required, and in this case addition is possible, and we are using the numbers cardinally. An illustration of this situation would be a set of arrows of appropriate lengths, pointing either to the left or to the right (see Figure 6(ii)). Experience with such an illustration makes it clear that the laws of addition for these numbers are that two numbers from the same set, positive or negative, add as if they were natural numbers, and the sum has the same sign as the component numbers; while addition of two numbers of different signs can be performed by the use of the law ${}^{+}a + {}^{-}a = 0$, together with the associative law, thus:

$$\begin{aligned} {}^{-}5 + {}^{+}3 &= ({}^{-}2 + {}^{-}3) + {}^{+}3 \\ &= {}^{-}2 + ({}^{-}3 + {}^{+}3) \\ &= {}^{-}2. \end{aligned}$$

Addition of two numbers of different sign is usually performed mentally by considering situations like those of Figure 7, or by using the subtraction theorem below, but we shall show that all these relationships may be made to depend on



${}^{-}5 + {}^{+}3 = {}^{-}2$ ${}^{-}5 + {}^{-}3 = {}^{-}8$ ${}^{+}5 + {}^{-}3 = {}^{+}2$

${}^{-}2 - {}^{-}5 = {}^{+}3$ ${}^{-}8 - {}^{-}5 = {}^{-}3$ ${}^{+}2 - {}^{+}5 = {}^{-}3$

Fig. 7 Addition and subtraction relationships for signed numbers.

the law $^{+}a + {}^{-}a = 0$, together with the associative, commutative and cancellation laws.

Subtraction, for signed numbers, cannot have the association with 'taking away' which is familiar in the case of natural numbers, since this idea derives from the characteristic application of natural numbers to sets of objects. Two arrows, or two changes of position on a scale, cannot be taken one from the other; these are the situations from which signed numbers have been derived, and to which we would wish any composition of them to apply.† However, the desirability of constructing a new number system which obeys laws as similar as possible to those of the old one leads us to define subtraction to have the same relation to addition in the new system as it had in the old: we say that, for integers α, β, $\alpha - \beta$ is a number which, added to β, gives α. The subtraction illustrations of Figure 7 may be obtained by direct application of this definition. We shall find that, as we proceed to consider further extensions of the number system, this principle will need constantly to be used—that the laws of composition of numbers in the new system are defined (i) so that they fit in with the old system in those parts which are common to both systems (in this case the positive integers) and (ii) so that the fundamental laws—associative, distributive, relationship of inverse and the rest—continue to be obeyed in the new system.

We may also derive from this definition the theorem of which we make constant use:

$$\alpha - \beta = \alpha + \beta',$$

where β' means the *opposite* of β, that is, the same natural number with the other sign (for example, $^{-}8 - {}^{-}5 = {}^{-}8 + {}^{+}5 = {}^{-}3$).

We now state our axioms for the integers and show how the properties we have discussed depend on them.

† By a *composition* of numbers we mean any way of associating a pair of them with a third number; addition, subtraction, multiplication and division are all compositions of numbers. We prefer this word to 'operation' since the latter is also used sometimes for transformations. We shall try to avoid it altogether.

AXIOMS FOR THE INTEGERS

The set J of *integers* consists of an element zero, together with the elements of two natural number sequences, one designated 'positive' and the other 'negative' (symbols $^{+}5$, $^{-}3$ are used), satisfying the following:

J1. Addition (+) and multiplication (.) are defined for integers, and they obey the associative, commutative, distributive and cancellation laws:

Associative laws:

1a. $\alpha + (\beta + \gamma) + \gamma, \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma.$

1b. $\alpha(\beta\gamma) = (\alpha\beta)\gamma.$

Commutative laws:

1c. $\alpha + \beta = \beta + \alpha,$ 1d. $\alpha\beta = \beta\alpha.$

Distributive law:

1e. $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma.$

Cancellation law:

1f. $\alpha + \beta = \alpha + \gamma \Rightarrow \beta = \gamma,$ 1g. $\alpha\beta = \alpha\gamma \Rightarrow \beta = \gamma.$

J2. Subtraction and division are defined from addition and multiplication as follows:

$\alpha - \beta$ is a number x such that $\beta + x = \alpha$ (if such a number exists);

$\alpha \div \beta$ is a number y such that $\beta . y = \alpha$ (if such a number exists).

J3. The positive and zero numbers combine in the same way as the natural numbers to which they correspond:

$$^{+}a + {}^{+}b = {}^{+}(a + b), \quad {}^{+}a . {}^{+}b = {}^{+}ab,$$

and similarly for zero.

J4. $^{+}a + {}^{-}a = 0.$

Definition The *opposite* α' of an integer α is the same natural number with the other of the two signs.

(In each section, we shall normally denote the new type of number being introduced by a Greek letter, and the old type by an ordinary Italic letter.)

Theorem 1 (a) ${}^-a + {}^-b = {}^-(a + b)$,

(b) ${}^+a + {}^-b = {}^-b + {}^+a = \begin{cases} {}^+(a-b) \text{ if } b < a, \\ {}^-(b-a) \text{ if } a < b. \end{cases}$

Proof (a)

$$\begin{aligned} ({}^-a + {}^-b) + ({}^+a + {}^+b) &= ({}^-a + {}^+a) + ({}^-b + {}^+b) && \text{by } J1 \\ &= 0 && \text{by } J4 \\ &= {}^-(a + b) + {}^+(a + b) && \text{by } J4. \end{aligned}$$

Thus

$${}^-a + {}^-b = {}^-(a + b) \qquad \text{by } J3 \text{ and } J1.$$

Proof (b) First,

$${}^+a + {}^-b = {}^-b + {}^+a \qquad \text{by } J1.$$

Then, if $b < a$,

$$\begin{aligned} {}^+a + {}^-b &= {}^+[(a - b) + b] + {}^-b && \text{by } J2 \\ &= {}^+(a - b) + ({}^+b + {}^-b) && \text{by } J1 \text{ and } J3 \\ &= {}^+(a - b) && \text{by } J4 \text{ and } J3. \end{aligned}$$

If $a < b$,

$${}^-b = {}^-a + {}^-(b - a) \text{ by } J2 \text{ and part (a),}$$

so

$$\begin{aligned} {}^+a + {}^-b &= {}^+a + [{}^-a + {}^-(b - a)] \\ &= {}^+(a + {}^-a) + {}^-(b - a) && \text{by } J1 \\ &= {}^-(b - a) && \text{by } J4 \text{ and } J3. \end{aligned}$$

Theorem 2 For all integers α,β, $\alpha - \beta = \alpha + \beta'$.

$$\begin{aligned} \textit{Proof} \qquad (\alpha + \beta') + \beta &= \alpha + (\beta' + \beta) \\ &= \alpha && \text{by } J4. \end{aligned}$$

Thus, from the definition in *I2*,

$$\alpha + \beta' = \alpha - \beta.$$

Exercises

1. Use Theorem 2 to show that $\alpha - \beta$ exists for all integers α, β; and use the cancellation law to show that it is unique (that is to say, that there cannot be two different integers satisfying the definition).

2. Define *greater* and *less* for integers.

3. Use the cancellation laws to show that $\alpha + 0 = \alpha$ and $\alpha.0 = 0$ for all integers α. (*J3* explicitly defines 0 only as a zero for positive integers.)

4. Show that there can only be one element with the properties of the zero.

5. Recast Theorem 1 to avoid the use of the commutative law $\alpha + \beta = \beta + \alpha$ *for integers*. (*J3* implies its truth for *positive* integers: use only this.) Hence *deduce* that $\alpha + \beta = \beta + \alpha$ for all integers (thus showing that it is unnecessary to include this as an axiom in *J1*).

MULTIPLICATION

The multiplication law ${}^-a.{}^-b = {}^+ab$ is difficult to derive on intuitive grounds, since there are few obvious applications of signed numbers in which multiplication has a meaning: temperature changes and bank deposits do not admit multiplication. Speed and time do, and one can illustrate the rule ${}^-a.{}^-b = {}^+ab$ from this situation, but the real reason for this and the other rules of signs lies in the need to produce a system in which the fundamental laws apply—in this case the commutative and distributive laws. Intuition may suggest that ${}^+2.{}^-3 = {}^-3 + {}^-3 = {}^-6$, making multiplication by a positive integer equivalent to repeated addition, and the commutative law then suggests that ${}^-3.{}^+2$ should also $= {}^-6$. But the product of two negatives cannot be deduced without appeal to the distributive law. The complete set of rules may be deduced quite briefly as follows:

Theorem 3 (a) For all integers α,β, $\alpha.\beta' = (\alpha\beta)'$.

(b) ${}^+a.{}^+b = {}^-a.{}^-b = {}^+ab$,
${}^+a.{}^-b = {}^-a.{}^+b = {}^-ab$.

Proof (a) ${}^+a.{}^+b = {}^+ab$ is given in *J3*. For the others, we have

$$\begin{aligned} \alpha.\beta' + \alpha.\beta &= \alpha(\beta' + \beta) && \text{by the distributive law, } J1\text{e},\\ &= \alpha.0 && \text{by } J4\\ &= 0\\ &= \alpha\beta + (\alpha\beta)' && \text{by } J4. \end{aligned}$$

Thus, by the cancellation law,

$$\alpha.\beta' = (\alpha\beta)'.$$

Proof (b) Putting $\alpha = {}^+a$, $\beta = {}^+b$ in this result, and using the commutative law, gives the second result of (b), while the first is obtained by putting $\alpha = {}^-a$, $\beta = {}^+b$ in the same relation and using the other result.

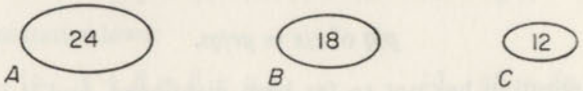
Exercises

1. 'I walk forward three paces: ${}^+3$; I turn round through 180°: $-$; I walk backward two paces: ${}^-2$; where am I now?' Which of the above theorems does this example illustrate? Find other illustrations of Theorems 2 and 3.

2. In this development, the $+$ and $-$ signs attached to the integers were required simply to distinguish one of the two component sets of natural numbers from the other: they had no connection with the use of the same signs for addition and subtraction. Investigate the consequences of changing *J3* so that the negative integers are the ones which combine like natural numbers, and consider whether this notation would be equally convenient in use.

THE RATIONAL NUMBERS

If the integers arise (at least in their cardinal form) when two natural numbers are compared by subtraction, as when two temperatures are compared to give a temperature change, the rational numbers may be considered to arise when two natural numbers are compared by division. The word *rational*, in fact, derives from *ratio*. If a set A contains 24 objects, and a set B has 18, we may say that B has 6 objects less than A, or $\frac{3}{4}$ as many objects as A. Composites of rational numbers occur readily in such situations; Figure 8 shows how an 'of' composition, $\frac{2}{3}$ of $\frac{3}{4}$, and an addition, $\frac{3}{4} + \frac{1}{2}$, may arise. It also shows a situation in which the Principle of Equality is used (that, for any integer k, the rational numbers p/q and kp/kq are equal) in giving the ratio of the numbers in B and A in its lowest terms. If we consider the addition here of the numbers of the sets B and C, and write in full:



B is $\frac{3}{4}$ of A C is $\frac{2}{3}$ of B
C is $\frac{2}{3}$ of $\frac{3}{4}$ of A
B and C together $= \frac{3}{4}$ of $A + \frac{1}{2}$ of $A = \frac{5}{4}$ of A

Fig. 8 Compositions of rational numbers.

$$\begin{aligned} B \text{ and } C \text{ contain } &\tfrac{3}{4} \text{ of } A + \tfrac{2}{3} \text{ of } B\\ &= \tfrac{3}{4} \text{ of } A + \tfrac{1}{2} \text{ of } A\\ &= \tfrac{3}{4} \text{ of } A + \tfrac{2}{4} \text{ of } A\\ &= \tfrac{5}{4} \text{ of } A, \end{aligned}$$

we may note that the rational numbers to be added must be expressing the quantities B and C both in terms of the same third quantity A, and that to perform the addition the denominators must be made equal by using the Principle of Equality, and then the numerators added. In symbols, the law is

$$\begin{aligned} p/q + r/s &= ps/qs + qr/qs\\ &= (ps + qr)/qs. \end{aligned}$$

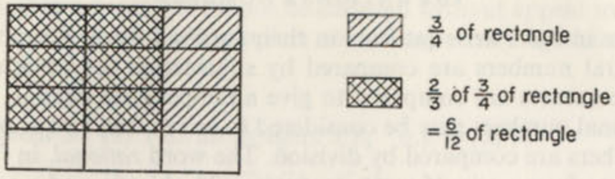

$\frac{3}{4}$ of rectangle
$\frac{2}{3}$ of $\frac{3}{4}$ of rectangle
$= \frac{6}{12}$ of rectangle

Fig. 9.

Figure 9 shows a simple way of illustrating the 'of' law,

$$p/q \text{ of } r/s = pr/qs.$$

Multiplication has not so far been mentioned, and it is not very obvious intuitively what meaning should be attached to this word for rational numbers. Multiplication of a rational by a natural number could be interpreted as repeated addition, for instance $3 \times \frac{2}{5}$ as $\frac{2}{5} + \frac{2}{5} + \frac{2}{5}, = \frac{6}{5}$: and the 'of' law above gives $\frac{3}{1}$ of $\frac{2}{5} = \frac{6}{5}$, so if the rationals with denominator 1 are to correspond to natural numbers, 'of' for rationals must correspond to multiplication for natural numbers. We shall use the symbol × for the 'of' composition in what follows. We shall also expect this composition to be distributive over addition. We now proceed to give a set of axioms for the rationals, and to show how the familiar properties may be deduced from them. We shall see that the axioms correspond closely to those for the integers, with the important addition of axiom *Q4*b.

The reason for this difference is that in the rationals we have not simply a double set of natural numbers or integers, but a set of *pairs* of these; not just the integers and the unit fractions $1/q$, but numbers p/q for any p, q. Another aspect of this difference is that we have whole sets of numbers like $\frac{3}{4}, \frac{6}{8}, \frac{9}{12}, \ldots$ which are equal: such a set is called an equivalence class.

The axioms we give here are for the whole set of rationals, positive and negative: we derive them from the integers.

AXIOMS FOR THE RATIONALS

Definition The set Q of rational numbers is a set of ordered pairs of integers (in which the second member of the pair is not zero), (symbol p/q, read 'p-by-q') satisfying the following axioms:

Q1. An equality relation, and laws of addition (+) and multiplication (× or ·) are defined for rationals, and they obey the associative, commutative, distributive and cancellation laws:

Associative laws:

1a. $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$, 1b. $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.

Commutative laws:

1c. $\alpha + \beta = \beta + \alpha$, 1d. $\alpha\beta = \beta\alpha$.

Distributive law:

1e. $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

Cancellation laws:

1f. $\alpha + \beta = \alpha + \gamma \Rightarrow \beta = \gamma$,

1g. $\alpha\beta = \alpha\gamma$ and $\alpha \neq 0 \Rightarrow \beta = \gamma$.

Q2. Subtraction and division are defined from addition and multiplication as follows:

$\alpha - \beta$ is a number x such that $\beta + x = \alpha$ (if such a number exists);

$\alpha \div \beta$ is a number y such that $\beta . y = \alpha$ (if such a number exists).

Q3. The rationals $p/1$, with second member 1, combine in the same way as the integers p.

$$p/1 + q/1 = (p + q)/1, \quad p/1 \times q/1 = pq/1;$$

0/1 and 1/1 have the properties of 0 and 1.

*Q4*a. $q/1 \times 1/q = 1/1$.

b. $p/q = p/1 \times 1/q$.

The Principle of Equality may be deduced from these axioms.

Theorem 1 (Principle of Equality)

If k is any integer,

$$kp/kq = p/q.$$

Proof We show first that $k/kq = 1/q$.

$$\begin{aligned} q/1 \times k/kq &= q/1 \times (k/1 \times 1/kq) && \text{by } Q4\text{b} \\ &= (q/1 \times k/1) \times 1/kq && \text{by } Q1\text{b} \\ &= qk/1 \times 1/kq && \text{by } Q3 \\ &= 1/1 && \text{by } Q4\text{a} \\ &= q/1 \times 1/q && \text{by } Q4\text{a.} \end{aligned}$$

Thus $k/kq = 1/q$ by $Q1$g.

Then

$$\begin{aligned} kp/kq &= kp/1 \times 1/kq && \text{by } Q4\text{b} \\ &= p/1 \times k/1 \times 1/kq && \text{by } Q3 \\ &= p/1 \times k/kq && \text{by } Q4\text{b} \\ &= p/1 \times 1/q && \text{proved} \\ &= p/q && \text{by } Q4\text{b.} \end{aligned}$$

Next we have a theorem (corresponding to Theorem 2 for integers) which shows that, with rational numbers, a division may always be replaced by a multiplication. We first define *reciprocal.*

Definition For any rational number $\alpha = p/q$, its *reciprocal* α^{-1} is the number q/p. (If $p = 0$ the reciprocal is not defined.)

Theorem 2 If α, β are rationals, and β^{-1} exists,

$$\alpha \div \beta = \alpha.\beta^{-1}.$$

Proof We show that $\alpha.\beta^{-1}$ satisfies the definition of $\alpha \div \beta$ in $Q2$. Let $\beta = p/q$. We have that

$$\begin{aligned} (\alpha\beta^{-1})\beta &= \alpha \times q/p \times p/q && \text{by } Q1\text{b} \\ &= \alpha \times q/1 \times 1/p \times p/1 \times 1/q && \text{by } Q4\text{b} \\ &= \alpha \times q/1 \times 1/1 \times 1/q && \text{by } Q4\text{a} \\ &= \alpha && \text{by } Q3 \text{ and } Q4\text{a} \\ &= (\alpha \div \beta).\beta && \text{by } Q2 \text{ and } Q1\text{d,} \end{aligned}$$

so $\alpha\beta^{-1} = \alpha \div \beta$ by $Q1$g.

We now prove, from the axioms and these theorems, the rules for addition and multiplication of rationals.

Theorem 3 (a) $p/r + q/r = (p + q)/r$,
(b) $p/q + r/s = (ps + rq)/qs$.

Proof (a)

$$\begin{aligned} p/r + q/r &= p/1 \times 1/r + q/1 \times 1/r && \text{by } Q4 \\ &= (p/1 + q/1) \times 1/r && \text{by } Q1\text{e and 1d} \\ &= (p + q)/1 \times 1/r && \text{by } Q3 \\ &= (p + q)/r && \text{by } Q4\text{b.} \end{aligned}$$

(b)

$$\begin{aligned} p/q + r/s &= ps/qs + qr/qs && \text{by Theorem 1} \\ &= (ps + qr)/qs && \text{by part (a).} \end{aligned}$$

Theorem 4 (a) $p/q \times q/r = p/r$,
(b) $p/q \times r/s = pr/qs$.

Proof (a)

$$\begin{aligned} p/q \times q/r &= p/1 \times 1/q \times q/1 \times 1/r && \text{by } Q4\text{b} \\ &= p/1 \times 1/1 \times 1/r && \text{by } Q4\text{a} \\ &= p/r && \text{by } Q3 \text{ and } Q4\text{b.} \end{aligned}$$

(b)

$$\begin{aligned} p/q \times r/s &= pr/qr \times qr/qs && \text{by Theorem 1} \\ &= pr/qs && \text{by part (a).} \end{aligned}$$

Exercises

1. Show that if two rationals y,y' are both values of $\alpha \div \beta$, then they must be equal.

2. Show that, since we have defined the rationals as pairs of integers, the cancellation law for addition follows from the remainder of the axioms.

3. Show that the above work can be recast so as to avoid stating either commutative law, for addition or multiplication, as an axiom; show that these can be proved from the remaining axioms, with some modifications; for instance, two distributive laws are required, one with the multiplier on the left of the bracket, one with it on the right.

POWERS AND ROOTS FOR INTEGERS AND RATIONALS

Repeated multiplication of a natural number by itself was defined above (p. 34) as a separate operation, called raising to a power. The same definition will serve for the raising of a rational to a power which is a natural number, but new definitions are required for negative and rational powers. We give these below.

Definition If a is a rational number and b is a natural number, a^b is defined by

$$a^1 = a$$

and

$$a^{b^+} = a^b.a.\dagger$$

Theorem $a^0 = 1.$
(The proof is as for $a \varepsilon N$, see p. 35).

Axiom

Q5. If a, b are rational, a^b is defined so that

(a) if b is of the form $p/1$, with p natural, $a^b = a^p$,
(b) for all $a, b, c \varepsilon Q$, $a^b.a^c = a^{b+c}$, and $(a^b)^c = a^{bc}$.

Theorem (1) If $a \varepsilon Q$ and $n \varepsilon N$,

$$a^{-n} = 1/a^n.$$

Proof

$$\begin{aligned} a^n \cdot 1/a^n &= 1 && \text{by } Q4\text{a} \\ &= a^0 && \text{proved} \\ &= a^n.a^{-n} && \text{by } Q5\text{b}, \\ \therefore \quad 1/a^n &= a^{-n} && \text{by the cancellation law } Q1\text{g}. \end{aligned}$$

Theorem (2) If $a \varepsilon Q$ and $q \varepsilon N$,

$$a^{1/q} = {}^q\!\sqrt{a}$$

where ${}^q\!\sqrt{a}$ is defined for rational a, just as for natural a, that is as any number x such that $x^q = a$.

Proof

$$\begin{aligned} (a^{1/q})^q &= a^1 && \text{by } Q5\text{b} \\ &= a. \end{aligned}$$

$\therefore$ $a^{1/q}$ satisfies the definition of ${}^q\!\sqrt{a}$ given above. (We do not assume here either the existence or the uniqueness of ${}^q\!\sqrt{a}$ for any given a or q.)

† b^+ denotes the successor of b (see p. 24); do not confuse with ${}^+b$.

RATIONAL NUMBERS AND DECIMALS

The decimal system, which extends the place-value idea to the non-integral part of a rational number, has a number of practical advantages, making calculations with rationals subject to the same procedures as those with integers. It has also some interesting theoretical aspects, which we shall explore now. Decimals, like integers, may be expressed in different bases: in base 2, 0·1 is $\frac{1}{2}$, 0·001 is $\frac{1}{8}$. In general, in base b, the value of the nth place after the point is $1/b^n$, and we may make the general definition following.

Definition The *decimal* $\pm\, A{\cdot}a_1\, a_2\, a_3 \ldots a_n$, in base b (where $a_1, a_2 \ldots$ are integers between 0 and $b-1$ inclusive), is the rational number $\pm\,(A + a_1/b + a_2/b^2 + a_3/b^3 + \ldots + a_n/b^n)$.

We shall use the term decimal even if the base is not ten, since there is no other term in common use.

Any given rational may be expressed as a decimal, provided we may choose the base, but if the base is given, as in the usual case with base ten, complications may arise. Thus $\frac{1}{3}$ is 0·1 in base 3, and 0·2 in base 6, but in base ten we have no exact expression. The division of 1 by 3, according to the usual method, gives 0·3333. . . , and at whatever point one stops, there is still a remainder. Of course, what we do in practice is to write $0{\cdot}\dot{3}$ and call it a repeating decimal, but this merely hides the far-reaching implications of the step which is taken. Is $0{\cdot}\dot{3}$ exactly equal to $\frac{1}{3}$, or is it an approximation? If it is exact, in what sense? All we can say is that the sequence of numbers

$$0{\cdot}3,\ 0{\cdot}33,\ 0{\cdot}333,\ 0{\cdot}3333, \ldots$$

approaches the value $\frac{1}{3}$ more and more closely, the farther we go along the sequence; that is, $\frac{1}{3}$ is the *limit* of the sequence of numbers given.

Definition The sequence $u_1, u_2, u_3, \ldots u_n, \ldots$ is said to have the limit l if, given any positive number h, however small, there is a value $N(h)$ of n such that all the terms of the sequence beyond the Nth lie within the interval $(l - h, l + h)$; or in symbols, for all $n > N(h)$, $l - h < u_n < l + h$.

The question arises whether all rational numbers may be expressed either as exact decimals or as the limits of repeating decimals. Can we be sure that the division will not continue indefinitely without ever repeating the figures of the quotient? This question may be answered by considering the remainders in the division. As an illustration we show in Figure 10 the division to obtain a decimal for $\frac{2}{7}$. Here the remainders at each successive stage are 2, 6, 4, 5, 1, 3, 2, . . . , and after this the cycle must repeat. We have in this cycle all the six different remainders which are possible in division by 7. In the general case of division by q, there are $q - 1$ possible different remainders, and therefore the first remainder must occur again in at most $q - 1$ more stages: so the decimal for p/q repeats, and the length of the repeating cycle is at most $q - 1$.

```
   0·2857142...
  ____________
7)2·00...
  1 4
  ___
    60
    56
    __
     40
     35
     __
      50
      49
      __
       10
        7
       __
        30
        28
        __
         20
```

Fig 10. The decimal for $\frac{2}{7}$.

Exercises

1. Show that the decimals for $\frac{1}{13}$ and $\frac{2}{13}$ repeat with a cycle of 6 figures, and that the twelve possible remainders are divided into two sets of six, one set appearing in each division. Investigate $\frac{3}{13}$, $\frac{4}{13}$ and draw any further conclusions you can.

2. Find which decimals repeat in base 2: observe the lengths of the cycles.

3. Similarly investigate decimals in base 7.

Repeating decimals have many fascinating properties which depend on group properties; on p. 89, Exercises 4 and 5, it will be suggested how these preliminary investigations may be continued.

REAL NUMBERS

We were able to show in the previous paragraph that the decimals for all rational numbers were repeating ones; we must now consider whether it is possible to have unending decimals which never repeat. The example 0·1010010001 . . . , in which there are 1, 2, 3, 4, . . . zeros between successive ones, shows that this is indeed possible, and that therefore numbers exist which are not included in the set of rational numbers. This striking fact was discovered by Pythagoras, though not quite in this form, and to him and his contemporaries it seemed to shake the foundations of all their mathematics. Even for us it is not easy to accept that we have here a class of numbers which cannot be derived in a precise way from the integers or rationals, except as the limits of infinite sequences. Nevertheless, this class of *real numbers*, as they are called, is a very important one, because they form a *complete* set of numbers, in the sense that every convergent sequence of real numbers has a limit which is a real number, and this is not true of rational numbers.

In defining the real numbers as unending decimals, since we wish to include all the rationals within the set, we must provide the terminating decimals with an unending sequence of zeros following the last non-zero digit. Also, we must emphasise that we are in this way defining a real number as an infinite sequence of rationals: the real number 0·1010010001 . . . is nothing more than the sequence

$$0{\cdot}1,\ 0{\cdot}10,\ 0{\cdot}101,\ 0{\cdot}1010,\ 0{\cdot}10100,\ 0{\cdot}101001, \ldots$$

of which the nth term consists of the first n places of the decimal. We cannot even define the real number as the *limit* of this sequence; there is no rational number which is the limit of this sequence, and until we have defined real numbers, we have no numbers which are not rational.

Definition A *real number* is a sequence of rational numbers represented by an unending decimal.

Sums and products of real numbers are defined in an obvious way as the sequences whose terms are the sums of products of corresponding terms of the sequences for the numbers being combined.

Thus 1·23426 . . . + 2·18321 . . . gives the sequence

3, 3·3, 3·41, 3·417, 3·4174, 3·41747, ...

Giving a precise definition to these compositions is complicated by the occurrence shown here, where the first decimal place is not given correctly until one has reached the third term of the sequence for the sum. The situation is worse in the case of products, as the reader may verify, but the difficulty is clearly one of detail rather than of general principle. A similar comment applies to the fact that the decimal for a real number is not unique, since, for example, 1·2999 . . . = 1·3000 . . .

We now try to give particular point to this introduction of real numbers, by showing that $\sqrt{2}$ does not exist as a rational but does exist as a real number.

Theorem There is no rational number whose square is 2.

Proof Assume that there is such a number, and that, expressed in lowest terms, it is p/q. Then

$$(p/q)^2 = 2,$$

so

$$p^2 = 2q^2.$$

Both sides of this equation are integers, and the right-hand side is divisible by 2: so, therefore, is the left-hand side, p^2. But if p^2 contains the factor 2, so must p; that is, $p = 2k$, with k an integer. Then

$$p^2 = 4k^2.$$

Thus

$$2q^2 = 4k^2,$$

and so

$$q^2 = 2k^2.$$

Since the right-hand side of this contains the factor 2, so does the left-hand side; that is to say, q^2 and hence q, is divisible by 2. Thus both p and q contain the factor 2, contrary to the original assumption that the fraction was already in its lowest terms. Thus the original hypothesis was false, and there is no rational whose square is 2.

Theorem There is a real number α whose square is 2.

Proof Since $1^2 = 1$, and $2^2 = 4$, the integral part of α is 1. To find the first decimal place, try $1{\cdot}1^2$, $1{\cdot}2^2$, . . .

We find

$$1{\cdot}4^2 = 1{\cdot}96, \quad 1{\cdot}5^2 = 2{\cdot}25,$$

so the first decimal figure is 4.

For the next figure we try $1{\cdot}41^2$, $1{\cdot}42^2$ . . . and find

$$1{\cdot}41^2 = 1{\cdot}9881, \quad 1{\cdot}42^2 = 2{\cdot}0164,$$

so that

$$\alpha \text{ is } 1{\cdot}41 \ldots$$

Any desired number of decimal figures may be obtained in the same way. Thus α is defined as an unending decimal, that is, as a real number.

THE 'NUMBER' OF RATIONAL AND REAL NUMBERS

The sense of shock which most people experience on discovering for the first time that the rational numbers do not include all numbers probably depends on the fact that the rationals, as well as being an infinite set, are also densely packed together. We may express this precisely by saying that between any two rationals, however close, there is another rational, and so, by implication, an infinite number of rationals. Between $\frac{1}{2}$ and $\frac{2}{3}$ that is, $\frac{3}{6}$ and $\frac{4}{6}$, we have $3\frac{1}{2}/6$, that is $\frac{7}{12}$: between p/q and r/s we have $\frac{1}{2}$ $(p/q + r/s)$. Yet, as we have shown, it is possible to have a sequence of rational numbers converging to a limiting point which is a precisely defined point among the rationals, but which does not correspond to any rational number. Another comparison between the rationals and the reals is given by the fact that the rationals may be counted, that is to say, they may be put into one-one correspondence with the set of natural numbers, while this

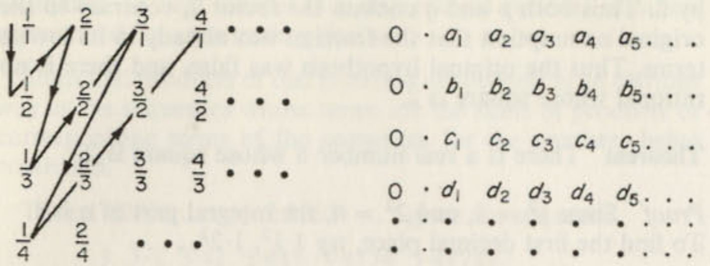

$\frac{1}{1}$ $\frac{2}{1}$ $\frac{3}{1}$ $\frac{4}{1}$ • • •
$\frac{1}{2}$ $\frac{2}{2}$ $\frac{3}{2}$ $\frac{4}{2}$ • • •
$\frac{1}{3}$ $\frac{2}{3}$ $\frac{3}{3}$ $\frac{4}{3}$ • • •
$\frac{1}{4}$ $\frac{2}{4}$ • • •
$0 \cdot a_1\ a_2\ a_3\ a_4\ a_5 \ldots$
$0 \cdot b_1\ b_2\ b_3\ b_4\ b_5 \ldots$
$0 \cdot c_1\ c_2\ c_3\ c_4\ c_5 \ldots$
$0 \cdot d_1\ d_2\ d_3\ d_4\ d_5 \ldots$

Fig. 11 The countability of the rationals and uncountability of the reals.

cannot be done for the reals. Figure 11 shows how the positive rationals may be put into an order and counted. Then if the reals are supposed to be capable of a similar ordering into a countable sequence, of which the first four members are shown, and if we form a new decimal by writing for the first figure any digit not equal to a_1, for the second figure any digit not equal to b_2, and so on, this decimal is one which cannot appear anywhere in the list, since it differs from every decimal in the list in at least one place. Thus the reals cannot be counted.

REFERENCES FOR CHAPTERS 2 AND 3

ADLER, I., *The New Mathematics*, Signet, New York, 1963.

GOODSTEIN, R. L., *Fundamental Concepts of Mathematics*, Pergamon, London, 1962.

THURSTON, H. A., *The Number System*, Blackie, London, and Interscience, New York, 1956.

4 COMPOSITIONS AND GROUP STRUCTURE

In the last chapter we were concerned chiefly with the extension of the natural number system to form the sets of integers and rational numbers. Historically the motive for both of these developments was probably a practical one—the new types of number were invented to describe situations to which the naturals did not apply. But the development may also be viewed as the abandonment of the chain-like structure of the naturals in favour of a group structure for addition and multiplication. The full meaning of this statement will be clearer when we have discussed the group concept in this chapter, but we may note at this stage that all the properties of natural numbers, including the ideas of addition and multiplication, arise from the fact that they form an ordered chain, with a starting point but no end. All the theorems in Chapter 2 depend on this fact, formulated as the Principle of Mathematical Induction—including even the Division Theorem, which would seem on the face of it to be very much about multiplication and addition, and hardly at all about order. But in proving properties of the integers and the rationals the Principle of Induction is of very little use, since although these sets can be made into an ordered chain, as shown at the end of the last chapter, this ordering does not correspond to the usual one in which $a < b$ means that $b - a$ is positive.

The essence of group structure is that the set contains an inverse for each element of the set, that is, in the case of addition, an element a' such that $a + a' = 0$, or for multiplication, an element $1/a$ such that $a \times 1/a = 1$. This makes possible the performance of an *inverse composition* (subtraction or division) within the set. The importance of the group concept, which we shall try to show in this book, is that certain sets which have nothing to do with numbers, notably sets of geometrical transformations and sets of permutations, also have these group properties, and the consequent interchange of ideas between these different fields is very productive.

COMPOSITIONS

We are already familiar with a number of ways in which two elements of a set may be combined to give a third. The elements 2 and 3 of the set J^+ of positive integers, for example, may be combined in the following ways, among others:

$$2 + 3 = 5, \quad 2.3 = 6, \quad 2^3 = 8, \quad 3^2 = 9,$$
$$2 - 3 = {}^{-}1, \quad 2 \div 3 = \tfrac{2}{3},$$
$$2\text{h}3 = 1 \text{ (meaning the hcf of 2 and 3 is 1).}$$

Any such rule by which every ordered pair (a, b) of elements of a set S is associated with a unique element c of the set is called an operation or *composition*$(*)$, on the set S, and we write $a \overline{*} b = c$. Note that $-$ and $\div$ are not compositions in this sense on the set J^+, since some pairs, like (2, 3) are associated by them with elements which are not in the set. We say in these cases that the set J^+ is not *closed* under subtraction or division.

We may also note in passing that the compositions $+$, $\cdot$ and h, are commutative and associative, while none of the other compositions used above has either of these properties. The power operation is not associative, since for example $(2^3)^4 = 2^{12}$, while $2^{(3^4)} = 2^{81}$.

Identity elements

Observe that, for all a in the set J^+,

$$0 + a = a + 0 = a, \quad 1.a = a.1 = a, \quad a^1 = a,$$

but $$1^a \neq a.$$

The element 0 is called an identity element for addition, since adding it to any element a leaves the element unchanged. Similarly 1 is an identity for multiplication. 1 also behaves like an identity for the power operation, but only if it is the right-hand element of the pair; we shall restrict the term 'identity element' to an element which leaves another unchanged when used on both the right and the left side of the other element. The h composition has no identity element.

Definition An element e of a set S is an *identity element* for a composition $*$ on S if $a * e = e * a = a$ for all a in S.

Inverse compositions and inverse elements

We are familiar with the close relationship between the compositions $+$ and $-$, and between $\times$ and $\div$, which are exemplified in such formulae as $(a + b) - b = a$, and in fact we have defined subtraction in terms of addition earlier in this book, by the statement that $a - b$ is x where $b + x = a$. But it is not easy to formulate a concise definition of an inverse composition from these facts; it is better to introduce the concept of *inverse element*. This enables us to dispense with such compositions as $-$ and $\div$, which is a considerable advantage since these compositions do not obey the commutative or associative laws. We have, in fact, already proved that, for any a and b in the set J, $a - b = a + b'$, and that for a, b in Q, $a \div b = a.(1/b)$. b' and $1/b$ are called the additive inverse of b and the multiplicative inverse of b, and in general we make the following definition:

Definition An element $\bar{a}$ is called an inverse of a with respect to the composition $*$ if $a * \bar{a} = \bar{a} * a = e$, where e is an identity element for the composition $*$.

(Sometimes an element is called a right-inverse if it satisfies the first requirement only, and a left-inverse if it satisfies the second only, but we shall not use these terms in this book.)

Exercises

1. Show that the hcf composition is associative.

2. Consider whether the following compositions in J^+ are commutative or associative, and whether an identity element and inverse elements exist.

   l, where $a l b$ is the lcm of a and b;

   min, where a min b is the smaller of a and b;

   max, where a max b is the greater of a and b.

3. If M is $\{1, 2, 3, \ldots 12\}$ and $(a \square b)$ is defined as the time b hours after a o'clock, is the set M closed under $\square$? If so, consider the same questions as in Exercise 2.

4. Formulate a definition of 'inverse composition', assuming inverse elements to be defined as in the text above.

5. If the 'power' composition p is defined by $apb = a^b$, consider whether either of the compositions q, r can be regarded as inverse to p, where $aqb = a^{1/b}$, and $arb = a^{\log_a b}$.

FUNCTION COMPOSITIONS

Almost all the sets which form groups, apart from number systems, consist of elements which can be classed as functions. Not all of these functions are like the familiar $y = x^2$ or $y = \sin x$ and it is convenient to adopt a modern definition of function which includes a wider class of mathematical objects.

Definition A function $f\colon X \rightarrow Y\colon x \rightarrow y$ is an association between the elements of a set X and a set Y such that each element x of X is associated with a unique element y of Y.

The chief difference between this and the older definitions of function is that here we take into account the set on which the function is defined, distinguishing between the function $y = x^2$ (or $x \rightarrow x^2$) defined for natural numbers and the same function defined for, say, the rationals. But what is really important in this new definition is that we regard the function as consisting not simply of a formula such as $x^2 + 3x - 2$, but of the whole set of ordered pairs (x, y) which are associated by it. In fact, an alternative definition is:

Definition A *function* $f\colon X \rightarrow Y\colon x \rightarrow y$ is a set of ordered pairs (x, y) such that each x of X is the first element of one and only one pair.

We shall say that *f maps X* into Y, or x onto y, and we shall use fx to denote the element of Y associated with the element x of X, calling it the *image* of x under the mapping.

The diagram of Figure 12 may help to clarify these ideas. The function consists of the set of pairs linked by the arrows.

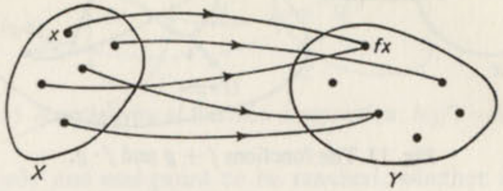


Fig. 12 A function.

There are three commonly-occurring ways in which functions can be combined, of which the third is the most important. The first two correspond closely to ordinary addition and multiplication.

Definition If functions f and g both map the set X into the set Y, and if an addition is defined in Y, then the sum $f + g$ is defined as the function mapping each x of X onto the element $fx + gx$ of Y.

Definition If functions f and g both map the set X into the set Y, and if a multiplication is defined in Y, then the product $f \cdot g$ is defined as the function mapping each x of X onto the element $fx \cdot gx$ of Y.

The third composition occurs when f maps X into Y, and g maps Y into a third set Z. We then have the possibility of an x being mapped by f onto its image fx in Y, and this element of Y being mapped onto its image $g(fx)$ in Z. Since each arrow of f has one and only one end-point in Y, and each point of Y is the start of an arrow which has one and only one end-point in Z, the mapping $x \rightarrow g(fx)$ is a function mapping X into Z. This combination of f and g is called *function composition*, and we denote the composite by $g \circ f$, or simply gf.

Definition If functions $f\colon X \rightarrow Y$ and $g\colon Y \rightarrow Z$ are defined, the *composite* (or *function-composite*), $g \circ f$ or gf, is defined as the function mapping each x of X onto the element $g(fx)$ of Z.

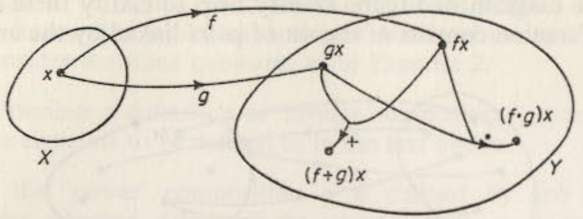


Fig. 13 The functions $f + g$ and $f \cdot g$.

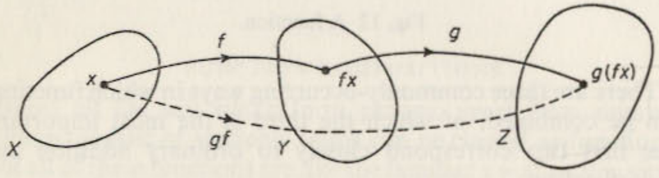


Fig. 14 The function gf.

Figures 13 and 14 illustrate these three compositions. gf is the 'function of a function' or 'substitution' composition, and occurs very frequently. It is, for example, the law of composition for permutations, and the law corresponding to matrix multiplication, and in this case is distributive over matrix addition.

The + and · compositions are associative and commutative, provided the corresponding compositions in Y are. Function composition is not necessarily commutative, as will be seen in the exercises below, but it is associative, and this fact is of far-reaching importance, since it enables a wide variety of functions to form groups.

Theorem Function composition is associative: $h(gf) = (hg)f$.

Proof The proof of this theorem is so simple that it is difficult to make it convincing. Figure 15 illustrates the situation, showing the sequence of arrows which starts from a single element x of X.

Each of the functions f, g, h, gf, hg consists of a whole set of arrows like the one shown. It is clear that if we start from x,

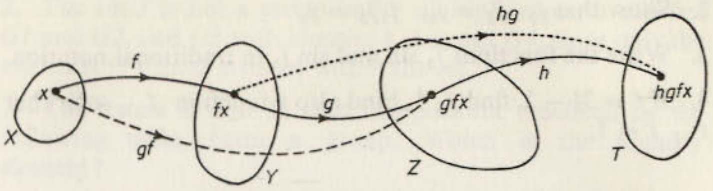


Fig. 15 Associativity of function composition: $h(gf) = (hg)f$.

there is only one end-point to be reached, whether we travel by f and then hg, or by gf and then h; and this would be true from whatever element x of X we started. We may write the composite function, without ambiguity, as hgf, and the image as $hgfx$.

Identity and Inverse Elements for Functions

Functions which act as identities for the + and · compositions may be readily found (see Exercise 1 below). The identity for function composition is the function which maps every element into itself: $x \to x$; we shall denote this function by the symbol j. Clearly $jj = j$, but $j \cdot j$ maps each x into x^2. We have not decided so far which of these to abbreviate as j^2: let us write j^2 for jj, and j_2 for $j \cdot j$. This gives us a convenient notation for the functions $x \to x^n$; we can denote them by j_n, and the polynomial function $x \to x^2 + 3x - 2$ becomes $j_2 + 3j - 2$. (We write 2 for the function $x \to 2$ for all x.) The inverse of a given function, as normally defined, is its inverse under function composition.

Definition The *inverse relation* f^{-1} of a given function f is the set consisting of the reversed pairs of f; it may or may not be a function.

From this it follows that if f^{-1} is a function, then $ff^{-1} = f^{-1}f = j$. A function whose inverse is also a function is called an *invertible function*.

Exercises

1. Show that the function 0, under which $x \to 0$ for all x, is an identity for the + composition. What is the identity for · ?

2. Show that $j_3 \cdot j_3 = j_6$, $j_3 j_3 = j_9$.

3. Write the functions j_3 sin and sin j_3 in traditional notation.

4. If $f = 3j - 2$, find f^{-1}. Find also a function f_{-1} such that $f_{-1}\, f = 1$.

GROUP STRUCTURE

The landmarks in the historical development of the number system were the extensions of the system of natural numbers to include first fractions and later signed numbers. Both of these extensions may be viewed as satisfying a need to perform an inverse composition, $\div$ or $-$, within the system, as well as the primary compositions of $+$ and $\cdot$. Systems in which a composition and its inverse can both be performed (or in which every element has an inverse element) are important in many branches of mathematics; such systems are called *groups*. We use inverse elements in preference to inverse compositions in defining this structure, for the reasons given above.

Definition A *group* $G, *$, is a set of elements, together with a composition $*$, satisfying the following four postulates:

G1. Closure: For all a, b in G, $a * b$ is in G.

G2. Associativity: For all a, b, c in G, $a*(b*c) = (a_* b)*c$.

G3. Identity: There is a unique element e in G with the property that $e * a = a * e = a$ for all a in G.

G4. Inverses: For each a in G there is a unique inverse element $\bar{a}$ such that $\bar{a} * a = a * \bar{a}$ in G.

Note that the composition is not required to be commutative.

Examples of Groups

1. The set J of integers (positive and negative and zero) is a group under addition: the identity is 0 and each element a has an inverse $\bar{a}$.

2. The set J is not a group under multiplication; it satisfies *G1* and *G2*, and *G3* with identity 1, but not *G4*, since only the elements ± 1 have inverses within the set.

3. The letters O and E, with composition $*$ defined by the following table, form a group. Which is the identity element?

$*$	O	E
O	E	O
E	O	E

(The entry in the O row and the E column is the element $O * E$.)

4. The structure $\{a + b\sqrt{3}\}, \cdot$, with a and b integers, is not a group; if a, b are rational and $0 + 0\sqrt{3}$ is excluded it is a group.

If $a,b \;\varepsilon\; J$, $\{a + b\sqrt{3}\}$ is closed under multiplication, since

$$(a + b\sqrt{3})(a' + b'\sqrt{3}) = (aa' + 3bb') + (ab' + a'b)\sqrt{3}$$

which is of the form $c + d\sqrt{3}$ with c and $a \;\varepsilon\; J$. There is an identity element $1 + 0\sqrt{3}$; but not all elements have inverses. For to make

$$(a + b\sqrt{3})(\bar{a} + \bar{b}\sqrt{3}) = 1 + 0.\sqrt{3}$$

we must have $a\bar{a} + 3b\bar{b} = 1$, and $a\bar{b} + \bar{a}b = 0$, and, for example, if $a = 2$, $b = 0$ the first of these gives $\bar{a} = \frac{1}{2} \not\varepsilon J$. If $a,b \;\varepsilon\; Q$, however, the second equation gives

$$\bar{b}/\bar{a} = -\, b/a \quad \text{(provided } a, \bar{a} \neq 0\text{)}$$

and the first then gives, on substitution for $\bar{b}$,

$$a\bar{a} + 3b\bar{a}(-\, b/a) = 1,$$

or

$$\bar{a}\left(a - \frac{3b^2}{a}\right) = 1$$

giving

$$\bar{a} = \frac{a}{a^2 - 3b^2}, \quad \bar{b} = \frac{-\,b}{a^2 - 3b^2},$$

provided $a^2 \neq 3b^2$ (since a and b are rational, $a^2 = 3b^2$ is impossible).

Thus we have, as the inverse of $a + b\sqrt{3}$,

$$\left(\frac{a}{a^2 - 3b^2} - \frac{b}{a^2 - 3b^2}\sqrt{3}\right)$$

(or $(1/3b)\sqrt{3}$, if $a = 0$) which is of the required form. Thus *G4* is satisfied; *G1* is dealt with as above. *G2* follows from the associativity of rational numbers.

Exercises

1. Show that the set Q of rational numbers (positive, negative and zero) is a group under addition; and that it is a group under multiplication if zero is excluded.

2. Show that the even integers form a group under addition, but the odd integers do not.

3. Show that the set of all integers of the form $3k$, with k integral, form a group under addition.

4. Show that the four numbers ± 1, $\pm i$, form a group under multiplication.

5. Show that the numbers 1, ω, ω^2, where ω is a complex cube root of 1, form a group under multiplication.

6. If R_α stands for a rotation of a body about a fixed axis through an angle α, and the operation $\circ$ means 'following' show that $R_\alpha \circ R_\beta = R_{\alpha+\beta}$. Show further that the set $\{R_\alpha\}$ for all real α with the operation $\circ$ forms a group, and state the identity element and the inverse of R_α.

7. Show that, with the same data, the set of four elements R_0, $R_{\pi/2}$, R_π, $R_{3\pi/2}$, with operation $\circ$, is a group.

8. Consider a situation in which there are two lamps, X and Y. There are four possible states: both off, X on, Y on, both on. There are also four possible *changes of state* which can be made: no change, change X, change Y, change both. Show that the four changes of state (denoted by N, X, Y, XY), with the operation 'following', form a group.

9. Show that the set of complex numbers of modulus 1 forms a group under multiplication.

10. Consider the set of all subsets of the set $\{a, b, c\}$; there are 8 such subsets, including the empty set and the set $\{a, b, c\}$ itself. Show that this forms a group under the composition $\triangle$, the *symmetric difference*, defined as follows: $A \triangle B$ is the set of all elements in A or in B but not in both.

11. Show that the set of all subsets of the set $\{a, b\}$, with the composition $\triangle$ defined in Exercise 10, forms a group which has a similar structure to that of Exercise 8.

12. Show that all rational numbers of the form 3^n (n an integer) form a group under multiplication.

13. Show that the rational numbers of the form $\dfrac{1 + 2m}{1 + 2n}$ with $m, n \varepsilon J$ form a group under multiplication.

14. Show that the set of matrices $\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$, taking all four possible combinations of the signs, forms a group under matrix multiplication.

15. Show that the set of four matrices $\begin{pmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{pmatrix}$ is not closed under multiplication, but together with the four matrices of Exercise 14 forms a group of eight elements.

16. Show that it is unnecessary to postulate the uniqueness of the identity element, or of the inverse of a given element in a group. Show, that is, that the uniqueness of these can be deduced from the remainder of the postulates.

Groups of Functions

We have shown above that the law of function composition ('function of a function') is associative. With this as the group operation, many sets of functions form groups, and these form a very significant class of groups, including, in fact, all symmetry and permutation groups. Most of these will be considered in detail later; we discuss here a few examples from different spheres.

The sets of functions: $x \to ax$, $x \to x + b$, $x \to ax + b$, defined for x in the rational field Q, and with a and b in Q $(a \neq 0)$, each form groups under function composition. Considering the general case f_{ab}: $x \to ax + b$, we have, for *G1*, $f_{ab} \circ f_{ab}$ is the function $x \to x''$ where $x' = ax + b$ and $x'' = cx' + d$, so that

$$\begin{aligned} x'' &= c(ax + b) + d \\ &= acx + (bc + d) \end{aligned}$$

which is of the required form.

For *G3* the identity is the function j: $x \to x$; and for *G4*, f^{-1}_{ab} is, from above, f_{cd} with $ac = 1$, $bc + d = 0$, which give $c = 1/a$, $d = -b/a$: the inverse is

$$x \to \frac{1}{a}x - \frac{b}{a}.$$

The set of all functions which map the set $\{1, 2, 3\}$ onto itself forms another group, called the group of permutations of 1, 2, 3 and denoted by P_3. The set of symmetries of a rectangle (that is, those rigid movements which leave the rectangle as a whole occupying the same position) are another group, which can be seen to be of the same type, for if the vertices are labelled 1, 2, 3, 4, any permissible movement is a permutation of these four symbols, with a restriction to those permutations which are geometrically possible. These do form a group; for two geometrically possible permutations performed successively are equivalent to a third which is also geometrically possible; and similarly for the inverse. These two types of group—permutation groups and symmetry groups— are discussed in Chapters 7 and 8 respectively.

Exercises

1. Show that the sets of functions $x \to ax$, and $x \to x + b$ $(x, a, b \in R)$ $(a \neq 0)$, form groups under function composition; find identities and inverses.

2. Show that the functions $x \to \dfrac{ax + b}{cx + d}$, $(x, a, b, c, d \in R)$ form a group under function composition, subject to a certain restriction on the values of a, b, c, d. State this restriction.

3. Consider whether the sets of functions (a) $x \to ax^2 + bx + c$ (b) $x \to ax^3 + bx + c$ form groups. Consider the x's and coefficients (i) in Q (ii) in R.

4. Show that the set of functions $(x, y) \to \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$, with $x, y, a, b, c, d \in R$, form a group, subject to a condition (to be found) on a, b, c, d.

5. Denoting the symmetry movements of a rectangle by h (reflection in a horizontal axis), v (reflection in a vertical axis), r (rotation through 180° in its own plane about its centre) and I (no movement), make up the group composition table. Is this group similar to any previously discovered ones?

6. Draw up the table of composition for the symmetries of the equilateral triangle.

Subgroup and Order of a Group

If a group $G,*$ contains a set H of elements which themselves form a group under $*$, the group $H, *$ is said to be a *subgroup* of the group G. The group of even integers under addition is a subgroup of the group of all integers under addition $(J, +)$; and the group $\{\pm 1, \pm \mathrm{i}\}, \cdot$ is a subgroup of the group of non-zero complex numbers under multiplication.

The *order of a group* is the number of elements it contains.

GENERATORS AND CYCLIC GROUPS

Consider the groups $\{\pm 1, \pm \mathrm{i}\}, \cdot$ and $\{N, X, Y, XY\}, \cdot$ of Exercises 4 and 8 on p. 64; form the 'powers' (with respect to the group operation) of one of the elements.

For instance, $\mathrm{i}^2 = -1, \mathrm{i}^3 = -\mathrm{i}, \mathrm{i}^4 = 1, \mathrm{i}^5 = \mathrm{i}, \mathrm{i}^6 = -1, \ldots$ and $Y^2 = N$, $Y^3 = Y$, $Y^4 = N$, $Y^5 = Y, \ldots$

In the first case the powers of i include all the elements of the group; we say that the element i *generates* the group. In the second case the element Y does not generate the group, and it may easily be verified (do this) that no other single element generates the group. This group is, however, generated by the *two* elements X and Y; since $X^2 = Y^2 = (XY)^2 = N$, all the products and powers of X and Y reduce to N, X, Y or XY.

A group which may be generated by a single element is called a *cyclic* group.

Every element a of a group G forms, together with all its different powers, a cyclic subgroup of G (which may be the whole group G). The order of this cyclic subgroup is called the *order of the element a*. Alternatively, the order of the element a may be defined as the smallest positive integer n for which a^n equals the indentity.

(If the group G is finite, the order of every element is a divisor of the order of the group. This is Lagrange's Theorem, which will be proved in Chapter 6.)

Exercises

Consider the groups described in the last two sets of exercises. In each case, find cyclic and non-cyclic subgroups (where possible) and sets of generators. Note the orders of the subgroups found, and compare them with the orders of the groups.

REFERENCES

ALEXANDROFF, P. S., *An Introduction to the Theory of Groups*, Blackie, London, 1959.

LEDERMANN, W. *The Theory of Finite Groups*, fourth edition, Oliver and Boyd, Edinburgh, 1964.

PAPY, G., *Groups*, Macmillan, London, and St. Martin's Press, New York, 1964.

5

RESIDUE CLASSES, GROUPS, RINGS AND FIELDS

CONGRUENCES

Consider the arrangement of the integers into six columns shown in Figure 16.

0	1	2	3	4	5
·	·	·	·	−8	−7
−6	−5	−4	−3	−2	−1
0	1	2	3	4	5
6	7	8	9	10	11
12	13	14	15	16	17
18	19	20	·	·	·

Fig. 16 Residue classes modulo 6.

Observe that any two numbers in the same column differ by a multiple of 6; and that any two numbers in the same column leave the same remainder on division by 6. We describe this by saying that numbers in the same column are *congruent modulo 6*.

Definition The integers a and b are said to be congruent modulo m, written $a \equiv b \pmod{m}$, if $a = b + km$, where k is an integer.

Exercises

1. Verify that $17 \equiv 2 \pmod{5}$, $-3 \equiv 15 \pmod{9}$.

2. Verify that 1, 10, 100, 1000 are all congruent to each other modulo 9.

3. State two integers congruent to 3 (modulo 11) and two congruent to -1 (modulo 4).

Let us label the columns of Figure 16, as shown, writing $\mathbf{0}$ to stand for the set $\{-6, 0, 6, \ldots\}$ and so on. We shall investigate this table further.

The columns of Figure 16 are called *residue classes modulo 6*, and we shall denote them by $\mathbf{0}_6, \mathbf{1}_6, \mathbf{2}_6, \ldots \mathbf{5}_6$; the suffix 6 will be omitted when the modulus is obvious from the context.

Exercises

1. Note the residue class modulo 6 to which each of the following integers belongs:

13, 20, 33; 7, 62, 69; 19, 8, 27;
8, 15, 23; 20, 33, 53; 8, 16, 24; 15, 10, 25.

Write a note about what you observe within each set of three numbers, and give a reason if possible.

2. Do the same with the following sets of integers:

8, 9, 72; -3, 10, -30; 3, 7, 21; 9, 5, 45.

Again make a conjecture about what you observe; make up some similar sets of integers to test your conjecture further; and try to prove it.

The above investigations lead to the following theorems:

Theorem If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then

$$a + c \equiv b + d \pmod{m},$$
$$a - c \equiv b - d \pmod{m}$$

and
$$ac \equiv bd \pmod{m}.$$

Proof We have $a = b + km$ and $c = d + k'm$ with k, k' integers; hence

$$a + c = b + d + (k + k')\,m,$$

and so
$$a + c \equiv b + d \pmod{m},$$

and similarly for subtraction.

Also
$$\begin{aligned} ac &= (b + km)(d + k'm) \\ &= bd + (kd + k'b + kk'm)\,m; \end{aligned}$$

thus
$$ac \equiv bd \pmod{m}.$$

Note that congruences cannot, in general, be divided: for $2 \equiv 8 \pmod{6}$, while $1 \not\equiv 4 \pmod{6}$. It can, however, be proved that *if c and m have no common divisor*, then

$$a \equiv b \pmod{m} \Rightarrow a/c \equiv b/c \pmod{m},$$

provided of course that a and b are both multiples of c. The proof of this is left as an exercise.

Applications of Congruences

Divisibility Tests It is well known that a number is divisible by 3 if the sum of its digits is divisible by 3, and by 4 if its last digit-pair is divisible by 4. All such tests can be discovered and proved by the use of congruences.

Example 1. To show that, in base ten, the number N, with digits *abcd* is divisible by 4 if the number with digits *cd* is divisible by 4.

We have

$$\begin{aligned} N &= d + c.10 + b.10^2 + a.10^3 \\ &\equiv d + c.10 + b.0 + a.0 \pmod{4} \\ &= \text{the number with digits } cd. \end{aligned}$$

Thus N and the number with digits *cd* leave the same remainder on division by 4, and hence if one is exactly divisible by 4, so is the other.

Example 2 To show that $N\ (= a_n a_{n-1} \ldots a_1 a_0)$, and the difference between the sums of the digits taken alternately,

$$(a_0 + a_2 + a_4 + \ldots) - (a_1 + a_3 + a_5 + \ldots),$$

leave the same remainder on division by 11; that is, that

$$N \equiv (a_0 + a_2 + \ldots) - (a_1 + a_3 \ldots) \pmod{11}.$$

We have $10 \equiv -1 \pmod{11}$ so that, by the theorem above,

$$10^2 \equiv (-1)^2 = 1 \pmod{11}$$

and $$10^3 \equiv 10^2.10 \equiv 1.-1 \equiv -1 \pmod{11}$$

and so on. So we have

$$\begin{aligned} N &= a_0 + a_1.10 + a_2.10^2 + \ldots a_n10^n \\ &\equiv a_0 + a_1(-1) + a_2.(1) \ldots + a_n(-1)^n \pmod{11} \\ &\equiv (a_0 + a_2 + \ldots) - (a_1 + a_3 + \ldots) \end{aligned}$$

as required.

Exercises

1. Prove tests for divisibility by 3, 9, 8, 99, 41 in base ten.

2. Find tests for divisibility by 5 and by 51_{six} ($= 31_{ten}$) in base 6.

3. Find tests for divisibility by various numbers, such as 11, 100, 101, 110, 111, . . . in base two.

Checks for Calculations Congruences may be used to check lengthy calculations, including those performed by machines.

Example 3 Check $8{,}297 \times 3{,}583 = 27{,}928{,}151$.

Check modulo 9:

$$\begin{aligned} 8{,}297 &\equiv 8 + 2 + 7 \equiv -1 \pmod{9}, \\ 3{,}583 &\equiv 3 + 5 + 8 + 3 \equiv 1 \pmod{9}, \\ 27{,}928{,}151 &\equiv 2 + 8 + 1 + 5 + 1 \equiv -1 \pmod{9}, \end{aligned}$$

and since $-1 \times 1 = -1$ the product given is either correct or differs from the correct product by a multiple of 9.

We will make a further check modulo 11:

$$\begin{aligned} 8{,}297 &\equiv (7 + 2) - (9 + 8) \equiv 3 \pmod{11}, \\ 3{,}583 &\equiv (3 + 5) - (8 + 3) \equiv 8 \pmod{11}, \\ 27{,}928{,}151 &\equiv 11 - 24 \equiv -2 \pmod{11}, \end{aligned}$$

but $$3 \times 8 = 24 \equiv +2 \pmod{11},$$

so there is an error.

(The correct product has the 7 and 9 interchanged; it is worth noting that this type of error is undetected by the modulo 9 check.)

Checks modulo 99 and modulo 101 are particularly valuable as they are easy to perform, and for an incorrect result to remain undetected by both tests it would need to differ from the correct result by a multiple of 99.101: an unlikely event.

Exercises

1. Check $7{,}342 \times 2{,}591 = 19{,}032{,}122$ modulo 9 and modulo 11.

2. Check $728{,}223 \times 5{,}535{,}064 = 4{,}030{,}760{,}911{,}272$ (a) modulo 9 and modulo 11; (b) modulo 99 and modulo 101. Comment.

3. Do some calculations on a machine and check them.

4. Prove that, if checks modulo a and modulo b are successfully applied to a calculation, the result is correct modulo ab.

RESIDUE CLASSES

We now return to the investigation of residue classes with which we began this chapter, and formulate the conclusions which might be drawn from the exercises on p. 70.

Theorem If **A** and **B** are residue classes modulo m, then all the integers c which may be obtained as the sum of an integer of **A** and an integer of **B** lie in the same residue class **C**. This class **C** is called the 'sum' of the classes **A** and **B**, and we write $\mathbf{C} = \mathbf{A} + \mathbf{B}$.

Similarly, there is a class **D** containing all products of an integer of **A** with one integer of **B**, and it is called the product of the residue classes **A** and **B**; $\mathbf{D} = \mathbf{A} \, . \, \mathbf{B}$.

Moreover both these compositions are associative and commutative, and multiplication is distributive over addition. The proofs are left as exercises; they follow directly from the basic theorem on congruences on pp. 70-71

Exercises

1. Prove the division of congruences, stated above (p. 71).

2. Prove the existence of a unique sum and product for two residue classes as stated above.

3. Prove that the compositions $+$ and $\cdot$ for residue classes are associative, commutative and distributive.

4. Verify that the addition and multiplication tables for the residue classes modulo 6 are as shown in Figure 17.

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

·	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Fig. 17 Addition and multiplication tables for residue classes modulo 6.

5. Verify the truth of the associative and distributive laws in some particular cases in the tables of Figure 17, for example $(4 + 2) + 3 = 4 + (2 + 3)$; $4.(2.3) = (4.2).3$; $4(2 + 3) = 4.2 + 4.3$.

Investigation

Consider how far the usual arithmetical operations can be performed in the system of Figure 17. Is subtraction possible? If $2 - 5$ is to be an element x such that $5 + x = 2$, is there such an x? Look at the table: we need to look along the row belonging to **5** and to find **2** . . . Answer, $x = 3$. Can *any* number be subtracted from any other?

Since $2 + 4 = 0$ we can say that **4** is the *additive inverse* of **2**, and write $^{-}2 = 4$; but we are not thereby introducing 'negative numbers' into the system, since the number we need, **4**, is there already. Find the additive inverses of the other elements of this system. Note that we can use these inverses for subtracting, since $3 - 4 = 3 + {}^{-}4 = 3 + 2 = 5$.

We also have powers; 3^4 can be defined as $3.3.3.3$, which is **3**. In fact $3^n = 3$ for every natural number *n*. Verify that $2^2 = 4$, $2^3 = 2$, $2^4 = 4^2 = 4$, $2^5 = 2$, and so on; and list the powers of **4** and **5**.

We cannot, however, divide any number by any other. If $2 \div 4$ is x, where $4.x = 2$, then the table gives $x = 2$ or **5**; but $2 \div 5$ does not exist, since $5.y$ is not **2** for any y.

We may conclude from this investigation that the set of residue classes modulo 6 forms a group under the composition $+$. It is not a group under $\cdot$. We shall denote this set of residue classes by the symbol J_6. The structure $J_6, +, \cdot$ is called the *finite arithmetic modulo 6*.

We next investigate the structure J_5. The tables are shown in Figure 18.

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Fig. 18 Tables for J_5.

Investigation

1. How does J_5 compare with J_6 with regard to addition? Can you find additive inverses for all elements?

2. Show that (a) the powers of **2** and **3** include all the non-zero elements, (b) the powers of **4** are all **4** or **1**, (c) $x^4 = 1$ for every element x except **0**.

3. Show that every element **a**, except zero, has a multiplicative inverse a^{-1} such that $a.a^{-1} = 1$, and hence that division is possible by any element except zero.

4. Solve a number of linear equations, such as $2x = 3$, $3x + 4 = 2$ and others; can every such equation be solved in J_5?

5. Try solving some quadratic equations, such as $x^2 = 2$, $x^2 + 4 = 2$, $x^2 + 3x + 2 = 0$, and others. Can every such equation be solved? What determines whether or not a quadratic can be solved in J_5?

6. Can all cubics of the form $x^3 = a$ be solved in J_5?

From this investigation we conclude that:

(i) The structure J_5, + is a group.

(ii) The structure J_5, · is not a group, but it becomes a group if the zero is excluded. We shall denote this by $J_5^4, \cdot$, the 4 indicating the number of elements remaining in the set. A structure like J_5 is called a *field*; J_6 is not a field but it is an example of a *ring*, a structure which is a group under addition and also has multiplication.

Further conclusions are:

(iii) All equations of the form $ax + b = c$ are soluble in J_5, provided $a \neq 0$.

(iv) **1** and **4** have square roots, but **2** and **3** do not.

This last may be compared with the situation in the system of real numbers. Just as the latter can be extended to make the system of complex numbers, so J_5 can be extended to include square roots for every element of the set. This will be considered in Chapter 9.

A Multiplicative Group from J_6

Although J_6 does not give a group under multiplication by the exclusion of **0** alone, a group can be obtained if the elements **0**, **2**, **3**, **4** are all excluded; this leaves just **1** and **5** with the multiplication table shown in Figure 19. It may be denoted by $J_6^2, \cdot$, and it is a group.

·	1	5
1	1	5
5	5	1

Fig. 19 The group $J_6^2, \cdot$.

Exercises

1. Write out the addition and multiplication tables for the finite arithmetic modulo 4 (J_4). Show that the elements **1**, **3**, form a group under multiplication (the group $J_4^2, \cdot$). Note the similarity in structure between this and the group $J_6^2, \cdot$.

2. Extract similarly the multiplicative groups from the arithmetics J_8, J_{10}, J_{12}. Show that they all contain 4 elements.

GROUPS, RINGS AND FIELDS

We now summarize the results of the above investigations, giving definitions and proofs.

Definition A *ring* $R, +, \cdot$ is a set R of elements, together with two compositions $+$, $\cdot$, satisfying the following axioms:

R1. R is a commutative group under the composition $+$.

R2. R is closed under the composition $\cdot$.

R3. The operation $\cdot$ is associative, and distributive over $+$.

Definition A *field* $F, +, \cdot$ is a set F of elements, together with two compositions $+$, $\cdot$, satisfying the following axioms:

F1. F is a commutative group under $+$.

F2. If 0 (the identity for $+$) is excluded, F is a commutative group under $\cdot$.

F3. $\cdot$ is distributive over $+$; that is to say, $a(b + c) = ab + ac$ for all a, b, c, in F.

Exercises

1. Show that the structure $J, +, \cdot$ (signed integers, with the usual addition and multiplication) is a ring.

2. Show that the set of all multiples of a given integer m, with $+$ and $\cdot$, is a ring (a subring of the ring $J, +, \cdot$).

3. Show that the structure of all polynomials with coefficients drawn from a ring R, is a ring.

4. Show that the structure Q of rational numbers is a field.

5. Show that the real numbers and the complex numbers also form fields.

6. Show that the set of numbers $a + b\sqrt{3}$, where a and b are rational numbers, is a field. State the zero, the unit, and the multiplicative inverse of $a + b\sqrt{3}$.

Theorem The set J_m $\{\mathbf{0}_m, \mathbf{1}_m, \ldots (\mathbf{m} - \mathbf{1})_m\}$, of residue classes modulo m, with the compositions of residue class addition and multiplication (see p. 73) forms a ring, called the *ring of residues modulo m.*

Proof The residue class compositions $\boldsymbol{+}$, $\cdot$, are associative, commutative and distributive ($\cdot$ over $\boldsymbol{+}$); this follows from the corresponding properties for the compositions $+$, $\cdot$ for integers. We give one proof in detail: the others are similar. Consider the classes $\mathbf{A} \,.\, (\mathbf{B} \boldsymbol{+} \mathbf{C})$ and $\mathbf{A} \,.\, \mathbf{B} + \mathbf{A} \,.\, \mathbf{C}$. The class $\mathbf{B} \boldsymbol{+} \mathbf{C}$ contains all integers of the form $b + c$, with $b \,\varepsilon\, \mathbf{B}$ and $c \,\varepsilon\, \mathbf{C}$. Thus $\mathbf{A} \,.\, (\mathbf{B} \boldsymbol{+} \mathbf{C})$ contains all integers of the form $a(b + c)$ $(a \,\varepsilon\, \mathbf{A})$. But these are precisely the integers of the form $ab + ac$, which constitute the set $\mathbf{A} \,.\, \mathbf{B} \boldsymbol{+} \mathbf{A} \,.\, \mathbf{C}$. Thus $\mathbf{A} \,.\, (\mathbf{B} \boldsymbol{+} \mathbf{C}) = \mathbf{A} \,.\, \mathbf{B} \boldsymbol{+} \mathbf{A} \,.\, \mathbf{C}$. It remains to prove that *R2* and the rest of *R1* are satisfied.

For *R2*: There is certainly a class $\mathbf{A} \,.\, \mathbf{B}$ containing the product ab of any two elements of $\mathbf{A}$ and $\mathbf{B}$ respectively; and this is unique, since if a_1, a_2 are any two integers of $\mathbf{A}$, and b_1,b_2 any two integers of $\mathbf{B}$, we have $a_1 \equiv a_2 \pmod{m}$ and $b_1 \equiv b_2 \pmod{m}$ so that $a_1b_1 \equiv a_2b_2$, by the Theorem, on p. 70 and hence a_1b_1 and a_2b_2 are in the same residue class.

For *R1*: (cf. *G1-4* on p. 62). The class $\mathbf{0}_m$ is an additive identity, since the class $\mathbf{0}_m + \mathbf{A}$ contains every element $a = 0 + a$ of $\mathbf{A}$, and so $\mathbf{0}_m + \mathbf{A} = \mathbf{A}$. And the class, $\bar{\mathbf{A}}$ say, which contains $m - a$, is the unique additive inverse of a, since $\mathbf{A} \boldsymbol{+} \bar{\mathbf{A}}$ contains the integer $(m - a) + a = m$, and so is the zero class $\mathbf{0}_m$. This completes the proof.

Theorem The ring of residues J_p, where p is prime, is a field. (It is sometimes called the Galois Field *GF(p)*.)

Proof We have already proved J_p to be a ring; it remains only

to show that, with zero excluded, J_p has an identity element and inverse elements for multiplication. It is immediately clear that the class $\mathbf{1}$ is a multiplicative identity. For inverses, we have to show that, for every class $\mathbf{A}$ except $\mathbf{0}$, there is a class $\mathbf{B}$ such that $\mathbf{A} \,.\, \mathbf{B} = \mathbf{1}$.

Consider the classes $\mathbf{A} \,.\, \mathbf{1}, \mathbf{A} \,.\, \mathbf{2}, \mathbf{A} \,.\, \mathbf{3}, \ldots \mathbf{A} \,.\, (\mathbf{p} - \mathbf{1}$;) these are the classes containing the integers $a \,.\, 1, a \,.\, 2, \ldots a(p - 1)$, where a is any integer of $\mathbf{A}$. These classes are all distinct, since no two of these integers are congruent modulo p; if they were, we should have

$$ar \equiv as \pmod{p}$$

and hence $\quad a(r - s) \equiv 0 \pmod{p}$,

that is $\quad a(r - s)$ is a multiple of p.

Thus either a or $r - s$ would need to be a multiple of p; this is impossible, since r and s are both between 1 and $p - 1$, and $a \,\varepsilon\, \mathbf{A}$, which is not the zero class.

Thus the $p - 1$ classes $\mathbf{A} \,.\, \mathbf{1}, \mathbf{A} \,.\, \mathbf{2}, \mathbf{A} \,.\, \mathbf{3}, \ldots \mathbf{A} \,.\, \mathbf{p} - \mathbf{1}$ are all distinct and so consist of the classes $\mathbf{1}, \mathbf{2}, \ldots \mathbf{p} - \mathbf{1}$ in some order. One of them is therefore $\mathbf{1}$, as required.

GENERAL PROPERTIES OF GROUPS, RINGS AND FIELDS

Much of the more developed theory of groups, with which we shall be concerned in subsequent chapters, applies only to finite groups. We include here those properties which apply to all groups, along with the most general properties of rings and fields.

Group Properties

Theorem: Cancellation Law In any group G, $*$,

$$a * b = a * c \Rightarrow b = c$$

and $\quad a * c = b * c \Rightarrow a = b.$

Proof $\quad a * b = a * c \Rightarrow \bar{a} * (a * b) = \bar{a} * (a * c) \quad$ by *G4*

$\Rightarrow e * b = e * c \quad$ by G2 and *G4*

$\Rightarrow b = c \quad$ by *G3*.

The proof of the other assertion is similar. Note that the cancellation law is true in some systems which are not groups, for example in the structure N, $+$ of natural numbers.

Theorem: Inverse of a Composite In any group G, $*$,

$$\overline{(a * b)} = \bar{b} * \bar{a}.$$

Proof
$$\begin{aligned}(\bar{b} * \bar{a}) * (a * b) &= \bar{b} * (\bar{a} * a) * b && \text{by } G2\\ &= \bar{b} * b && \text{by } G4\\ &= e && \text{by } G4,\end{aligned}$$

so that $\bar{b} * \bar{a}$ is a left-inverse for $a * b$.

Similarly $(a * b) * (\bar{b} * \bar{a}) = e$, so that $\bar{b} * \bar{a}$ is also a right-inverse, and so satisfies the requirements of $G4$ for an inverse of $a * b$. Moreover, in a group any inverse $\bar{c}$ of an element c is unique, since if $\bar{c}$ and $\bar{c}'$ are both inverses of c, we have $\bar{c} * c = e = \bar{c}' * c$, and by the cancellation law $\bar{c} = \bar{c}'$.

Theorem: Solution of Equations In any group G, $*$, the equations $a * x = b$ and $y * a = b$ have unique solutions.

Proof $a * x = b \Rightarrow \bar{a} * a * x = \bar{a} * b \Rightarrow x = \bar{a} * b$,

using $G2$, 4.

Similarly $\quad y * a = b \Rightarrow y = b * \bar{a}$.

Ring Properties

Theorem In any ring $R +, \cdot$, $a.0 = 0$ for all a.

Proof
$$\begin{aligned}a \,.\, a + a \,.\, 0 &= a\,(a + 0) && \text{by } R3\\ &= a \,.\, a && \text{since 0 is identity for } +\\ &= a \,.\, a + 0 && \text{since 0 is identity for } +,\end{aligned}$$

so by the cancellation law for $+$,

$$a \,.\, 0 = 0.$$

Exercises

1. Verify that the equalities $a^2 - b^2 = (a + b)\,(a - b)$ and $(a \pm b)^2 = a^2 \pm 2ab + b^2$ are true for elements in any commutative ring.

Field Properties

Theorem In any field F, $+, \cdot$,(i) the general linear equation $ax + b = c$ has a unique solution, provided $a \neq 0$; (ii) the general quadratic equation $ax^2 + bx + c = 0$, with $a \neq 0$, may be reduced to the form $(x + h)^2 = k$.

Proof We prove (ii), leaving (i) as an exercise.

$$ax^2 + bx + c = 0$$

$$\Rightarrow x^2 + \left(\frac{b}{a}\right) . \, x = -\frac{c}{a}$$

where $1/a$ is written for a^{-1},

$$\Rightarrow \left(x + \frac{b}{2a}\right)^2 = -\frac{c}{a} + \frac{b^2}{4a^2}$$

$$\Rightarrow \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

which is of the required form.

The reader should check that the manipulations required at each step are possible in any structure satisfying the field postulates. Note that the further solution of the equation requires that the element $(b^2 - 4\,ac)/4a^2$ should have a square root in the field F, and this is not so in general.

Exercise

Find which quadratic equations with coefficients in J_5 may be solved completely in J_5, and which cannot.

Notation for residue classes

When we are studying the properties of the structure J_m, $\boldsymbol{+}$, $\cdot$, the fact that its elements $\mathbf{0}_m$, $\mathbf{1}_m$, $\mathbf{2}_m$ . . . are actually residue classes is of no importance once it has been used to establish the composition tables for $\boldsymbol{+}$ and $\cdot$. What we are studying is simply a set of elements with certain composition relationships. We shall therefore abandon the former notation and denote these elements by the symbols 0, 1, 2 etc., and the compositions by ordinary + and · signs. It should be remembered, however, that the meaning of these symbols depends on the structure under consideration; for example, $3 + 3 = 1$ in J_5, + but $3 + 3 = 2$ in J_4, +.

ISOMORPHISM

Let us collect together for study the various groups of order 4 found in previous sections (p. 75 and p. 77 Exercise 2). The composition tables for these are shown in Figure 20.

+ mod 4	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

J_4, +

· mod 5	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

J_5^4, ·

· mod 8	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

J_8^4, ·

· mod 10	1	3	7	9
1	1	3	7	9
3	3	9	1	7
7	7	1	9	3
9	9	7	3	1

J_{10}^4, ·

· mod 12	1	5	7	11
1	1	5	7	11
5	5	1	11	7
7	7	11	1	5
11	11	7	5	1

J_{12}, ·

Fig. 20 Some residue groups of order 4.

Let us look for similarities among these tables. J_8^4 and J_{12}^4 show a close resemblance to each other. In fact, under the mapping

$$\begin{array}{cccc} 1 & 3 & 5 & 7 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 5 & 7 & 11 \end{array}$$

the first becomes exactly the same as the second. On inspection, it is clear that a similar relationship exists between J_5^4 and J_{10}^4, a possible mapping being $(1, 2, 3, 4) \to (1, 3, 7, 9)$. We have observed this resemblance visually, but it is clear that its basis lies in the fact that the structure of relationships among the elements of the two groups is the same; that is to say, that if $a \to A$, $b \to B$ and $c \to C$ under the mapping, and if $a * b = c$ in the first group, then $A * B = C$ in the second ($*$, $*$ being the compositions in the two groups). We call such a correspondence between two groups an *isomorphism*. Now consider Figure 21, which shows the table J_5^4 arranged with the elements in the order 1, 2, 4, 3. The visible similarity to J_{10}^4 has disappeared, but the structure of the group, that is the relationships among the elements, is unaltered. The isomorphism still exists although it is not visually obvious. Something else, however, has become visible—compare Figure 21 with the first table of Figure 20. We have an isomorphism between J_4, + and J_5^4, ·, with $(0, 1, 2, 3, +) \to (1, 2, 4, 3, \cdot)$.

· mod 5	1	2	4	3
1	1	2	4	3
2	2	4	3	1
4	4	3	1	2
3	3	1	2	4

Fig. 21 The group J_5^4, ·, rearranged.

We have now shown that among the groups of Figure 20, there are only two basically different structures. Can we perhaps show that these two are isomorphic, reducing to a single basic structure? So far our method of detecting isomorphisms has been by trial and error; we might try rearranging the positions of the rows and columns of the tables to see if we could observe further correspondences. In a later section we shall develop a more systematic method of seeking isomorphisms, using the generators of the groups; for the present it is sufficient to observe that J_{10}^4 and J_{12}^4 cannot be isomorphic, since (i) any isomorphism must have $1 \to 1$, and

(ii) the elements 3, 9 of J^4_{10}, whose squares are *not* 1, must → elements of J^4_{12} whose squares are not 1, and no such elements exist. Thus the two 4-element group structures we have found are essentially different.

It is useful to have ways of describing these two structures; this can be done by stating a set of generators and the relationships among them. One of these groups is cyclic, consisting of an identity 1 and elements a, a^2, a^3, with $a^4 = 1$; it is denoted by C_4. The other has elements 1, a, b, and ab, with $a^2 = b^2 = 1$ and $ab = ba$ (this implies $(ab)^2 = 1$); it is denoted by D_2, for reasons which will appear later, and is also known as Klein's four-group. The tables of these two 'abstract groups' are shown in Figure 22.

	1	a	a^2	a^3
1	1	a	a^2	a^3
a	a	a^2	a^3	1
a^2	a^2	a^3	1	a
a^3	a^3	1	a	a^2

C_4

	1	a	b	ab
1	1	a	b	ab
a	a	1	ab	b
b	b	ab	1	a
ab	ab	b	a	1

D_2

Fig. 22 The two abstract groups of order 4.

Definition An *isomorphism* between two structures S, $*$ and S', $*$ is a one-one mapping of S onto S' such that, for all a, b, c in S, if $a \to A$, $b \to B$ and $c \to C$, and if $a * b = c$, then $A * B = C$.

Definition An *automorphism* is an isomorphism of a structure onto itself.

Exercises

1. Show that (1, 3, 5, 7) → (1, 7, 11, 5) is another isomorphism between $J^4_8, \cdot$ and $J^4_{12}, \cdot$. Find a third. How many are there altogether?

2. Show that there are just two isomorphisms between $J^4_5, \cdot$ and $J^4_{10}, \cdot$.

3. State two isomorphisms between $J_4, +$ and $J^4_5, \cdot$.

4. Give values to a and b to show that $J^4_8, \cdot$ is an example of the abstract group D_2; show similarly that $J^4_{10}, \cdot$ and $J_4, +$ are examples of C_4.

5. Show that the mapping $(1, a, a^2, a^3) \to (1, a^3, a^2, a)$ is the only automorphism of the group C_4; find two automorphisms of D_2.

GENERATORS AND SUBGROUPS IN RESIDUE GROUPS

Example: Investigation of the group $J^6_7, \cdot$

The group has six elements 1, 2, . . . 6. We shall see what can be discovered without writing out the full composition table. We first seek generators.

$$2^2 = 4, \quad 2^3 = 4.2 = 1.$$

∴ 2 is of order 3.

$$3^2 = 2: \text{ so } 3^4 = 4,\ 3^6 = 1 \text{ while } 3^3 = 6, \quad 3^5 = 5,$$

and so 3 is a generator of the whole group.

Elements	1	2	3	4	5	6
	3^0	3^2	3^1	3^4	3^5	3^3
Orders	0	3	6	3	6	2

This is a cyclic group of order 6. All such groups are isomorphic, being represented by the abstract group C_6 with elements 1, a, a^2, a^3, a^4, a^5, with $a^6 = 1$.

The element 5, being of order 6, is also a generator of the whole group: we have:

1	2	3	4	5	6
5^0	5^4	5^5	5^2	5^1	5^3

The mapping $3^n \to 5^n$, that is (1, 2, 3, 4, 5, 6) → (1, 4, 5, 2, 3, 6), is the only non-identical automorphism of the group.

There are subgroups; $\{1, 3^3\}$ and $\{1, 3^2, 3^4\}$, that is $\{1, 6\}$ and $\{1, 2, 4\}$.

We note that the orders of these subgroups are 2 and 3, both divisors of 6.

Exercises

1. Make similar investigations of the groups J_9^6, $\cdot$, J_3, $+$, J_6, $+$, J_{14}^6, $\cdot$.

2. Show that, of the groups J_{15}^8, $\cdot$, J_{16}^8, $\cdot$, J_{24}^8, $\cdot$, two are isomorphic. Show that one of the structures represented is the abstract group generated by elements a, b, c with $a^2 = b^2 = c^2 = 1$ and each element commuting with each other ($ab = ba$, $ac = ca$ and so on). Give defining relations for the other abstract group.

3. Show that the groups J_m, $+$, for any m, are all cyclic and generated by the element 1.

6 COSETS AND LAGRANGE'S THEOREM

In the group J_7^6, $\cdot$, the set $\{1, 6\}$ forms a subgroup. The other four elements of the group may be divided into two sets of two elements each of which have a close connection with the subgroup; they are called *cosets* with respect to the subgroup $\{1, 6\}$ and are formed as follows. Any element of the group is taken, and is used to multiply each element of the subgroup; the resulting set is called the coset of that element with respect to the subgroup. Thus, in this case, the coset of the element 3 is the set $\{3.1, 3.6\}$, that is $\{3, 4\}$; and the coset of 4 is the set $\{4.1, 4.6\}$, which is $\{3, 4\}$ again. The coset of 2 is $\{2, 5\}$ and that of 5 is the same. The cosets of 1 and 6 are both the subgroup $\{1, 6\}$. The most important fact about cosets is immediately apparent here: two cosets (with respect to the same subgroup) are either the same, or else have no elements in common. This fact will have to be investigated later, and it is the basis of the proof of Lagrange's Theorem, but we shall not pursue it now.

Definition If H is a subgroup of a group G, and a is an element of G, the set of elements $a * h$, where h runs through every element of H, is called the left-coset of a with respect to H: it may be denoted by $a * H$. (The right-coset $H * a$ is defined similarly; in a commutative group the right- and left-cosets are the same.)

Exercises

1. Find the cosets in J_7^6, $\cdot$ with respect to the subgroup $\{1, 2, 4\}$.

2. Find the cosets in J_{11}^{10}, $\cdot$ with respect to the subgroup $\{1, 10\}$.

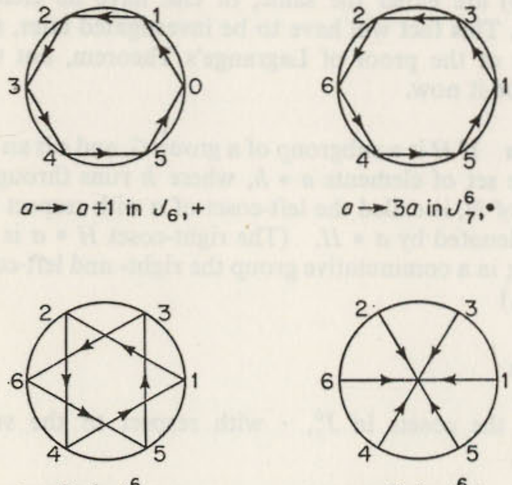
3. Find the cosets in J_6, + with respect to the subgroups $\{0, 3\}$ and $\{0, 2, 4\}$.

4. In Figure 24, verify that the cosets shown are correct, and obtain the right-cosets with respect to the subgroup $\{1, p\}$.

5. Find subgroups and cosets in the groups C_4 and D_2 (see Figure 22).

An Illustration of Cosets

Consider the second diagram of Figure 23. The elements of the group J_7^6, · have been allocated to the vertices of a suitable regular figure so that rotation through 60 degrees corresponds to multiplying by 3, in every case. To do this, we have only to express each element as a power of the generator 3 and number the vertices 1, 3, 3^2, 3^3, . . . in succession. Looking now at the third diagram, consider what happens if we start with 1, and multiply successively by 2. We go to 2, then to 4, and then back to 1; we have linked the numbers of the subgroup $\{1, 2, 4\}$ generated by 2. If we start at one of the other points, say 5, and multiply successively by 2, we go to 3, to 6, and then to 5 again: we have linked the numbers 5.1, 5.2, 5.2^2, that is, the numbers of the coset of 5 with respect to the subgroup $\{1, 2, 4\}$. Thus, in this figure, the subgroup and the coset are represented by two triangles, which we may perhaps describe as forming a 'broken' regular hexagon. The fourth diagram shows the result of multiplying successively by 6: the hexagon here breaks into three double-lines, representing the cosets with respect to $\{1, 6\}$. A similar investigation of the cosets in the additive group J_6, + may be performed using the first diagram of Figure 23.



$a \to a+1$ in J_6, + $a \to 3a$ in J_7^6, •

$a \to 2a$ in J_7^6, • $a \to 6a$ in J_7^6, •

Fig. 23 Regular polygons and cosets in some finite groups.

Exercises

1. Make similar diagrams for the groups J_8, + and J_{13}^{12}, ·, exhibiting the cosets as parts of the 'broken' regular polygons.

2. Investigate the important property of cosets referred to above, that they are either the same or have no common elements. Try to formulate reasons for this.

3. Show that the cosets in J, + with respect to the subgroup $\{3k\}$ are the residue classes $\{3k + 1\}$, $\{3k + 2\}$, $\{3k\}$ or $\mathbf{1}_3$, $\mathbf{2}_3$, $\mathbf{0}_3$.

4. Show that, in the division of 2 by 7 to give a repeating decimal, the remainders are congruent to 2, 2.10, 2.10^2, 2.10^3 and so on, mod 7, and so consist of the members of the coset of 2 with respect to the subgroup generated by 10 (that is, by 3) in the group J_7^6, ·.

5. Show that the subgroup generated by 10 in the group J_{13}^{12}, · is of order 6, and hence that the remainders in the divisions to find repeating decimals of the 13ths fall into two cosets of 6 elements each. Thus each of this set of decimals has one of two 6-digit cycles.

6. Show that, in the group of functions $R \to R$: $x \to ax + b$, for all $a,b \in R$, the set $x \to x + b$ forms a subgroup, and the coset with respect to this subgroup of a function $x \to Ax + B$, with fixed A,B, is the set $x \to Ax + b$, for fixed A and all possible b. If these functions are represented by their graphs

with respect to Cartesian axes, show that such a set of cosets is represented by a family of parallel lines of gradient A. Consider also the cosets with respect to the subgroup $x \to ax$.

LAGRANGE'S THEOREM

In this section we are inquiring into the relation between the numbers of elements in a group and in its subgroups, that is, between the *order* of a group and the order of its subgroups. We can only do this, of course, in the case of finite groups. The group C_4 of order 4, has one subgroup of order 2, $\{1,a^2\}$, while D_2, also of order 4, has three subgroups all of order 2, $\{1,a\}$, $\{1,b\}$, $\{1,ab\}$, (Figure 22). The isomorphic groups J_6, $+$ and J_7^6, $\cdot$, of order 6, have subgroups of orders 2 and 3. It appears that *the order of a subgroup is a divisor of the order of the group*, and this is in fact what Lagrange's Theorem states. A proof may be constructed using the properties of cosets with which we are now familiar: for we have seen that a subgroup of h elements divides the elements of the group into cosets, each containing h elements, so that every element of the group appears in one and only one coset. The order of the group must therefore be an integral multiple of h, the order of the subgroup. It remains only to prove that these assertions are true in general, and not only in the particular cases we have observed.

Theorem (Lagrange) If H, of order m, is a subgroup of a finite group G, of order n, then m divides n.

Proof Every left-coset with respect to H contains m distinct elements, for $a * h_1 = a * h_2 \Rightarrow h_1 = h_2$, so that each distinct member of H gives a distinct member of the coset: thus each coset contains m elements. Next, every element of G is in some coset: in fact, since the subgroup H must contain the identity e, the element $a = a * e$ is in the coset of a. Finally, no element can be in two different cosets: for if an element c is in the cosets of two elements a and b, we have $c = a * h_1$ and $c = b * h_2$ where h_1, h_2 are element of H. Thus $b = c * \overline{h_2} = a * h_1 * \overline{h_2} = a * h_3$, where h_3 is an element of H, showing that a and b are in the same coset, and therefore that the cosets of a and b are identical. Thus the n elements of G can be divided exactly into sets of m elements each: so m must be a divisor of n.

Direct Product Groups

Consider the set of ordered pairs (x, y), where $x \varepsilon Q, \cdot$ and $y \varepsilon J, +$. This set forms a group under the operation $*$ defined by $(x_1, y_1) * (x_2, y_2) = (x_1x_2, y_1 + y_2)$. The same applies if x and y are members of any groups, for example if $x \varepsilon \{1, a\}, \cdot$, with $a^2 = 1$, and $y \varepsilon \{1, b, b^2\}, \cdot$, with $b^3 = 1$. Such a group of pairs formed from two groups G, H is called the *direct product group* $G \times H$.

It may be proved generally (see Ledermann, p. 46) that a direct product group may be represented, not only as a set of pairs, but also as the group generated by the *products* of elements, one from each group, elements from the two groups being taken as independent of each other and so commuting.

Exercises

1. Show that the direct product group $C_2 \times C_3$ described above is isomorphic to the group generated by two independent, and therefore commuting, elements a, b, with $a^3 = b^2 = 1$; show also that this group contains elements of order 6, and thus that it is isomorphic to C_6.

2. Show that the group $C_2 \times C_2$ is isomorphic to the group D_2.

Possible Structures for Small Groups

In a previous section we examined a number of groups of order 4 and found them all isomorphic to one of two abstract groups, C_4 and D_2 (Tables, Figure 22). We also found it possible to describe these two groups completely by specifying the relations satisfied by their generators: C_4 is generated by a, with $a^4 = 1$, and D_2 by a and b, with $a^2 = b^2 = 1$, and $ab = ba$. We shall now show that these two are the only possible abstract groups of order 4, and consider similarly the possibilities for other orders.

First, the condition expressed in Lagrange's Theorem on the order of subgroups of a finite group, is at the same time a condition on the order of the elements of a group, since the order of an element is the order of the cyclic subgroup consisting of all the distinct powers of the element. Secondly, for any given order n there is at least one possible group—the cyclic group C_n, generated by a single element a with $a^n = 1$. Next, for a prime order p the cyclic group C_p is the only possible group, since Lagrange's Theorem requires that the order of every element shall be 1 or p, and any element except 1 will therefore generate the whole group.

This deals immediately with groups of orders 2 and 3, the only groups being $\{1, a\}$ with $a^2 = 1$, and $\{1, b, b^2\}$, with $b^3 = 1$: the groups C_2 and C_3. For order 4, we already know of two groups, C_4 and D_2, and we now consider whether any others are possible. Lagrange's Theorem tells us that the orders of the elements must be divisors of 4, that is 4 or 2. If there is any element of order 4, it will generate the group, which will thus be C_4. So to get a different group, all the elements except 1 must be of order 2. It is easy to show that the only possible group in this case is D_2; for, considering a generating set to include two elements a and b (there must be at least two to avoid a cyclic group), if we consider their product ab we have

$$ab = (ab)^{-1}$$

(since ab must be of order 2, and so equal to its inverse)

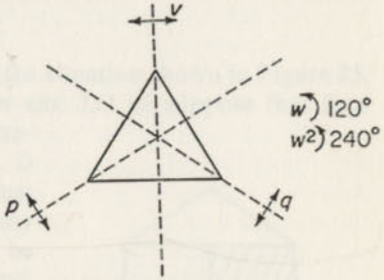
$$= b^{-1} a^{-1}$$
$$= ba$$

(since a and b, too, are equal to their inverses).

We have thus proved that any group of four elements, three of them being of order 2, contains an a, a b and an ab, and that $ab = ba$. This is sufficient to obtain the whole composition table and to see that it is identical with that of D_2.

For order 6, we have C_6, and any other group must have elements of orders 2 or 3 only. An extension of the argument of the previous paragraph shows that any group, all of whose elements except 1 are of order 2, is of order 2^n. (Since all the elements commute, each element is of the form $a^p b^q c^r \ldots$, where $p, q, r, \ldots$ are 0 or 1: the number of such elements, if there are n generators $a, b, c, \ldots$ is 2^n.) A group of order 6 must therefore have at least one element of order 3, a say, and at least one other element b. We therefore have as elements, $1, a, a^2, b, ba, ba^2, ab, a^2b, aba, bab$, and so on—all possible combinations of a and b —, and we have to find whether relations can be postulated among these elements which will reduce the number of distinct ones to six. The first six elements listed must be distinct, since, for instance, $ba^2 = a$ would imply $ba = 1$ and so $b = a^2$. Now consider b^2: this cannot be a or a^2, for similar reasons, and so must be 1. We now try to show how each element after the sixth in the list may be made equal to one of the six. For ab, we could have $ab = ba$ or $ab = ba^2$. The first of these gives $(ab)^2 = abab = a^2 b^2 = a^2 \neq 1$, and $(ab)^3 = a^2 ab = b \neq 1$, so that ab is not of order 2 or 3. (It is in fact of order 6 and we have the group C_6.) Taking the other choice, $ab = ba^2$, gives $(ba)^2 = baba = bba^2a = 1$: so ba is of order 2. We also deduce that $a^2 b = ba$, $aba = b$ and so on. It may be verified that this one non-cyclic group of order 6, which we may describe by the relations $a^3 = 1, b^2 = 1, (ab)^2 = 1$, is isomorphic to the triangle group D_3 whose table is given in Figure 24.

	1	w	w^2	v	p	q
1	1	w	w^2	v	p	q
w	w	w^2	1	q	v	p
w^2	w^2	1	w	p	q	v
v	v	p	q	1	w	w^2
p	p	q	v	w^2	1	w
q	q	v	p	w	w^2	1



Some left cosets $\{1\ w\ w^2\}, \{v\ p\ q\}$

$\{1\ p\}, \{w\ v\}, \{w^2 q\}$

Fig. 24 The symmetry group of the equilateral triangle.

For order 8, it may be proved by similar methods that there are five possible groups, three commutative ones, C_8, $C_4 \times C_2$,

$C_2 \times C_2 \times C_2$, and two others, the symmetry group of the square, D_4, defined by $a^4 = b^2 = 1$, $ba = a^3b$, and the so-called quaternion group defined by $a^4 = 1$, $a^2 = b^2$, $ba = a^3b$. (Ledermann, p. 51.)

Exercises

1. Show that the group generated by the following matrices is the quaternion group.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad \begin{pmatrix} 0 & -i \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

2. Show that the symmetry group of the square has the defining relations stated above.

3. Identify each of the groups J_{20}^8, J_{24}^8, and J_{30}^8 with the appropriate one of the five abstract groups described above.

4. Show that there are two groups of order 9, C_9 and $C_3 \times C_3$.

7

PERMUTATIONS

Group theory had its origin, not in the study of number systems or of the symmetry of geometrical figures, but in the course of investigations into the celebrated problem of the insolubility of the general equation of the fifth degree by means of formulae involving radicals. Galois showed that the solubility of equations depended on considerations of symmetry of functions of the roots such as $x_1 x_2 + x_2 x_3 + x_1 x_3$. This expression is unchanged in value if we make any interchange of the suffixes 1, 2, 3 among themselves, so that its 'symmetry' is closely related to that of the equilateral triangle. But the importance of permutations in group theory rests not only on this historical fact, but also on the fact that any finite group whatever is isomorphic to some permutation group, so that by studying permutation groups, we are studying all possible types of finite group.

Let us begin by considering the situation shown in Figure 25, which represents a four-seater car. Let us suppose that four people, A, B, C and D are travelling in the car, of whom D is the only driver and so must occupy seat 4. In how many ways can the other three be seated? . . . If A sits down first, he has the choice of 3 seats: then, whatever choice A has made, B has 2 seats to choose from, and after that C has Hobson's choice. So there are 3 × 2 or 6 different arrangements of A, B and C on the three seats. Let us describe the arrangement

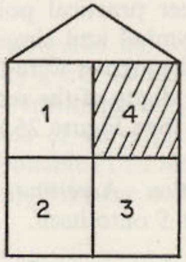


Fig. 25.

in which A is on seat 2, B on seat 1, and C on seat 3, by the symbol (2, 1, 3). Let us now go on to consider what changes can be made in this situation. Suppose A and C change places: we could denote this by $(A \rightleftarrows C)$, or by $(3 \rightleftarrows 2)$ meaning that the people in seats 2 and 3 have changed places. The resulting state of A, B, C is (3, 1, 2), and since we have decided to describe the seating-state in this way it will be convenient if we describe the changes, too, in terms of the seats rather than the people, that is to say, by $(3 \rightleftarrows 2)$. We may write

$$(2, 1, 3) \xrightarrow{(3 \rightleftarrows 2)} (3, 1, 2)$$

and this describes fully what has happened. There are just two other ways in which two people may change places, $(1 \rightleftarrows 3)$ and $(1 \rightleftarrows 2)$. If we now suppose that all three people wish to move at the same time, what changes are possible? . . . . One possibility is that the person in seat 1 moves to seat 2, the person in seat 2 goes to seat 3, and whoever was in seat 3 moves round to occupy the seat 1. Let us describe this by the symbol $(\overleftarrow{1 \rightarrow 2 \rightarrow 3})$. What other three-person movements are possible? We could have $(\overleftarrow{2 \rightarrow 3 \rightarrow 1})$: and perhaps $(\overleftarrow{3 \rightarrow 2 \rightarrow 1})$? . . . No, this is the same as $(\overleftarrow{1 \rightarrow 3 \rightarrow 2})$ or as $(\overleftarrow{2 \rightarrow 1 \rightarrow 3})$. So we have altogether five possible changes of seating in this situation—which is what we should expect, since we agreed that there were six different arrangements altogether. In practice we shall find it convenient to make the five changes up to six by including 'no change' and writing it as (1). Another practical point is that we can omit the arrows in this symbol and simply agree that the symbol is to be read as if the arrows were there, Thus the six possible interchanges of the digits of the set {1, 2, 3} are (1), (12), (13), (23), (123), (132). (See Figure 26.)

Definition A *permutation* of a set S is a one-one mapping of the set S onto itself.

The changes considered above are therefore the permutations of the set {1, 2, 3}: the notation (23), (132) and so on is called the *cycle* notation.

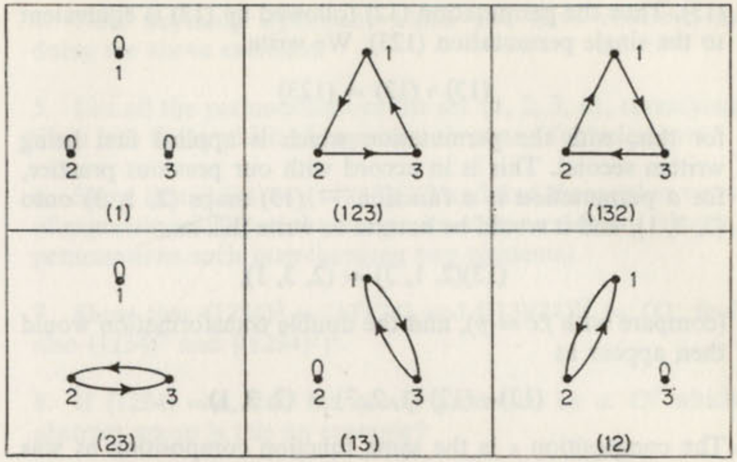


Fig. 26 The six permutations of P_3.

Exercise

Express the following permutations in cycle notation:
(3, 2, 1) → (1, 2, 3), (2, 3, 1) → (2, 1, 3), (3, 1, 2) → (2, 3, 1).

Suppose now that the occupants of the car wish for some reason to make the change

$$(1, 2, 3) \xrightarrow{(123)} (2, 3, 1)$$

(that is, starting from the state (1, 2, 3) to make the change (123), resulting in the final state (2, 3, 1)), but to do it, if possible, by a succession of interchanges involving only two people at a time, that is by permutations of the type (13). (These are called *transpositions*.) Is this possible? . . . Since 1 must → 2, we might start by applying (12). We have

$$(1, 2, 3) \xrightarrow{(12)} (2, 1, 3)$$

which can be made into the (2, 3, 1) that we want by applying

(13). Thus the permutation (12) followed by (13) is equivalent to the single permutation (123). We write

$$(13) \circ (12) = (123)$$

for this, with the permutation which is applied first being written second. This is in accord with our previous practice, for a permutation is a function – (13) maps (2, 1, 3) onto (2, 3, 1), and it would be natural to write this as

$$(13)(2, 1, 3) = (2, 3, 1),$$

(compare with $fx = y$), and the double transformation would then appear as

$$(13) \circ (12)\,(1, 2, 3) = (2, 3, 1).$$

The composition $\circ$ is the same function composition as was defined in Chapter 4. and just as the $\circ$ was often omitted there, so it will be here.

The method of calculating the composite of two or more permutations in the cycle notation is illustrated below.

To find $(13) \circ (23) \circ (13)$. Choose any digit, say 1, and begin a cycle thus: (1 . . .). What does 1 become? Working from right to left along the three permutations, 1 first becomes 3, then the 3 becomes 2 and the 2 is unaffected by the remaining permutation. So we will fill in (12 . . .). Now what does 2 become? Unchanged, then 3, then 1. So 2 goes to 1, and the cycle closes: (12). Now we must see what happens to 3. It goes to 1, then unchanged, then to 3. So we either write (12)(3), or we may leave out the 3 altogether:

$$(13) \circ (23) \circ (13) = (12).$$

Exercises

(In the following, $(12) \circ (13)$ will be written as (12)(13), the $\circ$ being omitted.)

1. Show that (23)(12) = (132) and find (12)(23).

2. Find (13)(12), (12)(13), (31)(32) and (32)(31).

3. Find (12)(12) (written $(12)^2$), $(23)^2$, $(123)^2$, $(123)^3$.

4. Note anything significant which you have observed in doing the above exercises.

5. List all the permutations of the set $\{1, 2, 3, 4\}$, classifying them into those which change two, three and four elements.

6. Show that (1234) = (14)(13)(12) and find three other ways of expressing (1234) as the composite of transpositions (that is, permutations each interchanging two elements).

7. Show that $(1234)^2 = (13)(24)$ and $[(13)(24)]^2 = (1)$; find also $(1234)^3$ and $[(1234)^3]^2$.

8. If $(1234) = a$, find the group generated by a. Of which abstract group is this an example?

9. Show that (1234)(4321) = (1). Is it true in general that $(abc \ldots)(\ldots cba) = (1)$?

PERMUTATIONS AND GROUPS

The above exercises show that certain sets of permutations may form groups, the composition being function composition, that is to say, one permutation being followed by the other.

We now prove a theorem about this.

Theorem The set of all permutations of the set $\{1, 2, 3, \ldots n\}$, under function composition, forms a group of order $n!$, denoted by P_n. (It is sometimes called the symmetric group of degree n, and denoted by S_n.)

Proof We verify that the group postulates *G1–4* are satisfied. It is clear from the definition of a permutation as a one-one mapping of a set onto itself that the result of two such mappings performed in succession is itself a permutation of the set. Next, we have already proved that the composition of functions is associative. There is an identity permutation (1), mapping each element onto itself; and to a permutation $(1, 2, 3, \ldots n) \rightarrow (a, b, c, \ldots g)$ there is the inverse $(a, b, c, \ldots g) \rightarrow (1, 2, 3, \ldots n)$.

Exercises

1. Show that the four permutations (1), (13), (24), (13)(24) form a group isomorphic to the group D_2.

2. Show that the three permutations (1), (123), (132) form a group isomorphic to the group C_3.

3. Show that the six permutations of three symbols discussed on p. 96 form a group (the group P_3), which is isomorphic to the group D_3.

4. Exercise 8 of the previous set shows that there is a group of permutations of 4 symbols (a subgroup of P_4) which is isomorphic to C_4. Find two further subgroups of P_4 which are also isomorphic to C_4.

5. Find a subgroup of P_4 which is isomorphic to D_4, by considering the symmetries of a square. (It is best to number the four corners of a square 'hole' as in Figure 39; then a movement in which the vertex of the inner square which was at 2 goes to 4 gives 2 → 4 as an element of the corresponding permutation. Thus (13) corresponds to the reflection in the axis 24. This corresponds to the way in which the permutation group P_3 is related to the seating situation discussed above.) Show that this subgroup may be generated by the permutations (1234) and (13).

6. Show that there are two other subgroups of P_4 which are isomorphic to D_4, corresponding to the other ways of assigning the symbols 1, 2, 3, 4, to the corners of the hole. Do these three 8-element subgroups contain all the 24 permutations of P_4? If not, which elements of P_4 do not appear at all?

7. Show that the inverse of (12)(1234) is (4321)(12).

8. Show the relation between the group of symmetries of the rectangle and a suitable subgroup of P_4.

9. Show how to obtain a group of permutations isomorphic to any cyclic group C_n.

PERMUTATIONS AND CYCLES

A number of points concerning the relations between permutations, their group properties, and the cycle notation have emerged from the above sets of exercises; we shall now formulate and prove these.

Theorem Every permutation may be expressed as a composite of *disjoint* cycles (cycles with no symbol in common).

Proof We may start with any element a of the set being permuted and build up a cycle in the normal way, as described on p. 98; thus $(afk\ldots)$. We either reach a eventually, thus closing the cycle, or we repeat some other previous element of the cycle, since we have only a finite set of symbols at our disposal. But we cannot reach an element without first having its predecessor in the cycle: we cannot, for example, have k without first having had f. Thus we cannot repeat any other element of the cycle without first repeating a: so the cycle must close in the normal way. If the cycle closes before all the symbols are exhausted, a new cycle may be started with one of the remaining symbols, and the same arguments apply to it. Moreover, no element from a previous cycle can appear since this would require the presence of its predecessor in the original cycle, and so on, so that we would need to have the whole of the original cycle, and this could only happen if we had started the new cycle with one of the previously used elements.

Note that if cycles are disjoint, as these are, they may be applied in any order without affecting the result: the composition of disjoint cycles is commutative. Also, there is only one way of allocating the elements of a given permutation to disjoint cycles.

Theorem The order of a cycle is equal to its length; the order of any permutation is equal to the lcm of the lengths of its disjoint cycles.

Proof If a one-cycle permutation $(abcd\ldots)$ of length n is applied once, $a \to b$: twice, $a \to c$: three times, $a \to d$. To make $a \rightarrow a$ the permutation must be applied n times: and then every element goes into itself. For the general permutation this must happen one or more times in all cycles of the permutation: the total number of applications required is the lcm of the lengths.

Theorem Every cycle $(abcd \dots k)$ of length n may be expressed as the composite of the n–1 transpositions $(ak) \dots (ad)(ac)(ab)$ (cf Exercise 6, p. 99). (The order of composition of these transpositions is significant, since they are not disjoint: and other expressions are possible.) Every permutation may thus be expressed as a composite of transpositions; and if it consists of p disjoint cycles involving a total of q elements, the number of transpositions obtained by this method is q–p.

Proof The composite of $(ak) \dots (ad)\,(ac)\,(ab)$ (working from right to left) is readily calculated to be $(abcd \dots k)$. The last part of the theorem follows immediately from the fact that each cycle of length n gives n–1 transpositions.

These transpositions are not in general disjoint, and the expression of a given permutation by transpositions cannot therefore be unique: we can, for example add the pair $(ab)(ab)$ to any such expression without altering the permutation. It is, however, an important fact that we can only change the number of transpositions by an even number. We can thus speak of an *even permutation*, as one which is always expressed by an even number of transpositions, or of an *odd permutation*. The proof that this alternating character, as it is called, of a permutation is constant will be investigated in the following exercises.

Exercises

1. Show that (14)(13)(12)(23)(34) = (1243) and express this as the product of just three transpositions.

2. Calculate the composite of the same set of cycles as in Exercise 1, taken in the reverse order, and express this, too, by three transpositions.

3. Find an expression for the composite (1342)(2145) in terms of six transpositions in one way, and in terms of four transpositions in two ways.

4. Show that (12)(13)(16)(26)(34)(13)(53) can be reduced to disjoint cycles by a repetition of the following process: wherever two cycles with common elements appear together, combine them to form a single cycle. At each stage of this process, calculate the number q–p where q is the number of the symbols involved (counting repeated ones each time they appear) and p is the number of cycles: show that q–p changes only by multiples of 2.

5. Combine the following pairs of cycles to form disjoint cycles:

$$(1ab2cde)(1fgh2jk),$$

$$(1a2b3c)(1p2q3r),$$

$$(1a2b3c4d)(1p2q3r4s),$$

where $a, b, c, \dots s$ are all distinct. Compare these and note what you observe. Show that q–p changes by 0 or 2.

6. Explain how, in Exercise 5, the one- and two-cycle results arise.

Theorem If any one representation of a given permutation by transpositions contains an even number of these, then every such expression contains an even number of them; and similarly for a permutation expressed by an odd number of transpositions.

Proof The unique expression of a permutation by disjoint cycles gives rise to an expression with a definite number of transpositions, as described in the previous theorem. Any other expression by transpositions may be reduced, as described in Exercise 4 above, to disjoint cycles; and in this process the number q–p changes only by a multiple of 2 at each stage. Exercise 5 shows how, if two cycles having an odd number of elements in common are combined, the result is a single cycle, whereas two cycles having an even number of elements in common reduce to two disjoint cycles; in both cases, all the repetitions disappear. Thus q–p decreases by an even number in both cases. But before the reduction, if the number of transpositions was k, then q–p was $2k$–k, that is, k: and in the final expression by disjoint cycles we have by the theorem above that the corresponding number of transpositions is equal to the final value of q–p. Thus the numbers of transpositions before and after reduction can differ only by a multiple of 2.

We now state the following theorem, which is an immediate consequence of the above.

Theorem The composite of two even or two odd permutations is even; the composite of one even and one odd permutation is odd.

To prove this, we have only to consider the permutations involved to be expressed by transpositions.

Exercises

1. Note that *single cycles* are odd if their length (and so their order) is even, and vice versa: but, for example, though (13) and (24) are odd, (13)(24) is even.

2. Show that the group P_3 contains three odd and three even permutations.

3. Show that, in the group D_4, the set of even permutations forms a subgroup; and identify this subgroup.

4. Show that the square of every permutation is even.

5. Show that in any permutation group containing both odd and even permutations, the even permutations form a subgroup.

6. Party-game: three tumblers, two of them upright, one upside down, are on the table. A 'move' consists of inverting any two of them. The problem is to get all three upright. Show, by consideration of odd and even permutations, that this is impossible.

The Alternating Group

Some of the examples have already suggested that if a group contains both odd and even permutations, exactly half are of each type. We now prove this. Consider what happens when all the permutations of a group containing both types are combined with a certain odd permutation, d. All the even ones become odd, and all the odd ones even. But since the composite of any permutation with d is another permutation of the group, and no two of the permutations obtained can be the same (since $dp_1 = dp_2 \Rightarrow p_1 = p_2$), what we are left with is still the same total set of permutations, and so there must be equal numbers of even and odd ones.

Definition The subgroup of P_n consisting of all its even permutations is called the alternating group of degree n, and denoted by A_n.

Exercises

1. List the permutations of A_3 and A_4.

2. Show that (3456) = (13)(1456)(13), and hence show how to express both this and any other permutation by transpositions of the form $(1a)(1b) \ldots (1k)$ even if the symbol 1 does not appear in the permutation.

3. Show that $(1a)(1b) = (1ba)$; hence express each of the permutations (12)(34), (13)(24), (14)(23) as composites of cycles of length 3.

4. Is a similar expression possible for the permutation (1234)?

5. Show that the group A_n is generated by the cycles (123), (124), . . . $(12n)$. Generate A_4 in this way.

CAYLEY'S THEOREM

In our investigations in this chapter, we have found a group of permutations isomorphic to every group we had previously found, and more besides. This is the reason for the importance of permutation groups: the fact is expressed in the following theorem.

Theorem (Cayley) Every group of order n is isomorphic to a subgroup of the permutation group P_n.

We use the following tables to illustrate the proof.

	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

	1	r	r^2	r^3
1	1	r	r^2	r^3
r	r	r^2	r^3	1
r^2	r^2	r^3	1	r
r^3	r^3	1	r	r^2

	1	a	b	c	$\dots$
1	1	a	b	c	$\dots$
a	a	a^2	ab	ab	$\dots$
b	b	ba	b^2	bc	$\dots$
c	c	ca	cb	c^2	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

These tables are of D_2, C_4, and a general group. The permutation to be associated with an element a is that which takes the symbols at the heads of the columns into the symbols of the row belonging to a; thus in the general table above, a is to be associated with the permutation $(1, a, b, c, \dots) \rightarrow (a, a^2, ab, ac, \dots)$ which may be described as the permutation Pa, under which, for each element x of the group, $x \rightarrow ax$. This *is* a permutation of the group elements: no two of ab, ac and so on can be the same since in a group $ax = ay \Rightarrow x = y$ (p. 79). The same group property also ensures that no two of these permutations is the same—in fact no two can have the same effect on any element: $bx = cx$ would Impiy $b = c$.

We have thus proved that to each element a of a group G there corresponds a permutation Pa of the group elements defined by $x \rightarrow ax$: and that this correspondence is one to one. To show finally that it preserves compositions, consider $Pa \circ Pb$. This first takes each x of G into bx: and as x runs through the elements of G, so does bx, so we can describe the effect of the following permutation Pa by $bx \rightarrow a(bx)$. Thus $Pa \circ Pb$ is the permutation under which $x \rightarrow abx$, that is, the permutation Pab, which proves the isomorphism, since it shows that the composition of the *permutations Pa, Pb* corresponds to the composition of the *group elements a, b*.

Exercises

1. Express in cycle notation the permutations associated with the elements of the group D_2, as tabulated above, and verify that they form a group isomorphic to D_2.

2. Do the same for the group C_4, and for another chosen group.

3. Represent as permutation groups some of the groups mentioned in the Exercises on pp. 64-5.

AUTOMORPHISMS

The reader may have noticed, in seeking isomorphisms between the groups of residues in Chapter 5, that some groups had non-trivial isomorphisms with themselves. For example the group $J_5^4, \cdot$ has the isomorphism $(1, 2, 3, 4) \rightarrow (1, 3, 2, 4)$: this may be shown either by writing out the two tables or by observing that the elements may be expressed both as $1, 2, 2^2, 2^3$, and as $1, 3, 3^2, 3^3$, and the correspondence is then $2^n \rightarrow 3^n$. Such an isomorphism of the set with itself is called an *automorphism*. Since under any isomorphism elements of the same order must correspond to each other, generators must correspond to generators, and so the group $J_5^4, \cdot$, having only 2 and 3 as generators, cannot have any more automorphisms except for the identity. We note that its automorphisms are (1) and the permutation (23). Now consider the group D_2 (Tables on p. 84). In this case any permutation whatever of the three elements a, b, c is an automorphism. These permutations form a *group* isomorphic to P_3: those of the previous case formed an example of C_2. Is it perhaps the case that the set of automorphisms of any group themselves form a group? It is easy to see that this is so: for an automorphism is simply a permutation of the group elements which preserves the group composition: if two such permutations are performed successively, the result is another permutation which preserves the group composition: thus the set of automorphisms of a group is closed under function composition. Similarly, this composition is associative, and the inverse of an automorphism is an automorphism. So the set of automorphisms does form a group.

Example To show that the automorphism group of C_7 is isomorphic to C_6.

In the group C_7 every element except the identity is of order 7, and may be taken as a generator of the group. Thus there are 6 possible automorphisms, mapping a given element a onto each of the elements a, a^2, a^3, a^4, a^5, a^6, in turn. For example, the automorphism which replaces the generator a by a^2 takes $(a, a^2, a^3, a^4, a^5, a^6)$ into $(a^2, a^4, a^6, a, a^3, a^5)$. Writing this as a permutation of the indices, it is (124)(365). Similarly if a maps onto a^3 the full permutation is $(1, 2, 3, 4, 5, 6) \rightarrow (3, 6, 2, 5, 1, 4)$, or, in cycle form, (132645). As the whole group may be generated by a single element, the choice of the element onto which a is to map determines the maps of all the elements: so having considered all the possibilities for a, there are no more: there are exactly 6 automorphisms. Moreover, we need go no further, for we have found one automorphism which is of order 6, so the automorphism group can only be C_6.

Exercises

1. Show that the automorphism group of C_5 is C_4, and that of C_6 is C_2.

2. Show that the automorphism group of D_3 is D_3, and find that of D_4.

3. Find the automorphism group of A_4.

CONJUGATE ELEMENTS

If any element of a group G is combined with another element of G to form the composite tat^{-1}, we say that a is 'transformed by t' and that a and tat^{-1} are conjugate elements with respect to G. This is an important relationship in several branches of mathematics. We shall investigate it here as it applies to permutations. For a simple case, take a as (1234) and t as (16). Then tat^{-1} is (16)(1234)(16), which is (6234): the 6 has replaced the 1 in the cycle. Next take a as (1435)(26) and t as (123). Then tat^{-1} is (123)(1435)(26)(132), which reduces to (1524)(36) or by rearranging, (2415)(36). This contains cycles of the same length as a, and again the permutation (123) has been applied to the symbols of a.

Let us see whether we can find a transforming permutation to take (123)(45) into (243)(15): the change of symbols required is (124) so we try this on the left and its inverse (142) on the right: (124)(123)(45)(142). It may easily be verified that this works as required.

It thus appears that two permutations are conjugate if and only if their disjoint cycles are the same in number and length. This and the other conclusions of the previous paragraph can be proved easily from the following theorem.

Theorem If f is a function mapping a set X into itself, and t is an invertible function mapping X onto a set Y, then the function g defined on Y which corresponds to f on X, that is which maps each element tx onto the element $t(fx)$, is given by $g = tft^{-1}$.

Figure 27 supplies the proof; we need only note that since t is invertible, every y in Y is of the form tx for some x in X, and then

$$tft^{-1}(tx) = tfx$$

as required (using the associativity of function composition).

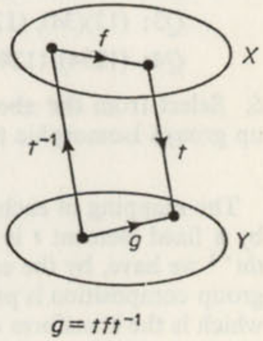


Fig. 27.

Applied to the permutation situation this theorem states that if a permutation t is applied to the symbols of the sets on which the permutation is defined, the corresponding permutation of the new set is tft^{-1}. Thus to obtain tft^{-1} we have only to apply t to the symbols of the cycles. It follows immediately that two permutations are conjugate, that is, related as f and tft^{-1}, if and only if their disjoint cycles are the same in number and length.

Exercises

1. Find all the conjugates of each element of the group D_3 (see p. 93).

2. Find the conjugates of each element of the group D_2: prove that in a commutative group each element is conjugate only to itself.

3. Find all the conjugates of each element of D_4.

4. Show that the conjugacy relation divides a group into a number of disjoint classes.

5. Show that the conjugacy classes of P_4 are as follows:

$Q0$: (1).

$Q1$: (12), (13), (14), (24), (34), (23).

$Q2$: (234), (243), (134), (143), (124), (142), (123), (132).

$Q3$: (12)(34), (13)(24), (14)(23).

$Q4$: (1234), (1243), (1324), (1342), (1423), (1432).

6. Select from the above list the elements of A_4: also make up groups isomorphic to C_4, D_2, D_3.

The mapping of each element of a group onto its transform by a fixed element t is an invertible function: for if $tat^{-1} = tbt^{-1}$ we have, by the cancellation laws, $a = b$. Moreover, the group composition is preserved, for $(tat^{-1})(tbt^{-1}) = t(ab)t^{-1}$ which is the transform of ab. It is therefore an automorphism. An automorphism which can be obtained in this way is called an inner automorphism; other automorphisms are outer.

Exercises

1. Find the inner automorphisms of the groups quoted in the exercises on p. 108.

2. Show that the inner automorphisms of a group themselves form a group, and identify this group in each of the cases in Exercise 1.

REFERENCES FOR CHAPTERS 4 TO 7

ALEXANDROFF, P. S., *An Introduction to the Theory of Groups*, Blackie, London, 1959.

LEDERMANN, W., *The Theory of Finite Groups*, fourth edition, Oliver and Boyd, Edinburgh, 1964.

PAPY, G., *Groups*, Macmillan, London, and St. Martin's Press, New York, 1964.

BIRKHOFF, G. and S. MACLANE, *A Survey of Modern Algebra*, Macmillan, New York, 2nd edn, 1953.

8

SYMMETRY GROUPS

An important field of application of group theory is the study of symmetry—the symmetry of natural and constructed objects, geometrical figures and of algebraic forms. In this chapter we shall begin with a brief study of algebraic symmetry, devoting the rest of the chapter to geometrical figures.

ALGEBRAIC SYMMETRY

The expressions $ab + bc + ca$ and $(a - b)(b - c)(c - a)$ may both be described as symmetrical, but they do not both possess the same degree of symmetry. The symmetry of each can be described by quoting the group of permutations of its letters which leaves the expression unchanged. In the first expression, any two letters can be interchanged without altering the expression, or, of course, a cyclic interchange can be made. The symmetry group is therefore the group P_3. In the second case, a transposition such as (ab) changes the sign of the expression, so a pair of such changes is required to leave it unchanged. In this case the group is the alternating group A_3, consisting of the even permutations of P_3.

Exercises

1. Show that the expression $(a - b)(b - c)(c - d)(d - a)$ has the symmetry group C_4. Include further brackets to form an expression with group A_4. State some expressions with group P_4.

2. Verify that the expressions giving the area and circumradius of a triangle ABC in terms of the sides and angles have the symmetry group P_3. Check similarly the symmetry of the expressions for the coordinates of the centroid, orthocentre,

incentre and circumcentre in terms of the coordinates of the vertices.

3. Show that the expression $(a - c)(b - d)/(a - d)(b - c)$ (called the cross-ratio of a, c, b, d) has a symmetry group isomorphic to D_2.

GEOMETRIC SYMMETRY

The set of letters and figures of Figure 28 exhibits a number of types of symmetry, including the familiar reflective symmetry of the A, B and H and the 'half-turn' symmetry of the N and the parallelogram. It is convenient to define a symmetry of an object as any *isometry* (that is, any distance-preserving transformation) of the points of the plane or of space which leaves the object as a whole unchanged. Then the full set of symmetries of the H consists of reflections in two perpendicular axes and a half-turn, together with the identity; and the symmetries of the swastika consist of the identity and rotations through one, two and three quarter-turns. It is immediately apparent that the *number* of symmetries does not by itself describe adequately the type of symmetry which the figure possesses, since both the H and the swastika have four. We can, however, distinguish these types by describing the way in which the symmetry transformations of each figure combine with each other, and it is at this point that group theory becomes relevant. If we denote by h the transformation which reflects every point of the plane containing the H in a horizontal axis, by v the similar reflection in the vertical axis, and by r the half-turn, it is easy to see that the composite of v and h, in either order, is r. (In Figure 29 $v \circ h$ makes $1 \to 4 \to 3$, which would

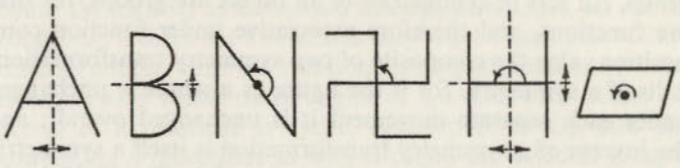
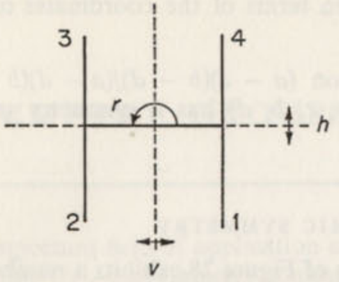


Fig. 28 Symmetries of some common figures.



$\circ$	I	h	v	r
I	I	h	v	r
h	h	I	r	v
v	v	r	I	h
r	r	v	h	I

Fig. 29 The letter H and its symmetry group.

be accomplished in one move by r.) The complete composition table is shown in Figure 29 and it is clear that we have here a *group*. All sets of symmetries of an object are groups, for they are functions, and therefore associative under function composition; also the composite of two symmetry transformations is itself a symmetry, for if the figure as a whole is unchanged under each separate movement it is unchanged overall: and the inverse of a symmetry transformation is itself a symmetry.

The group whose table we have found in Figure 29 is the group D_2 which we have met before: it is the symmetry group of the rectangle and the rhombus. Similarly the group of the swastika, r, r^2, r^3 and the identity $I\,(= r^4)$, is an example of the group C_4.

Exercises

1. Identify the groups of symmetries of the other figures of Figure 28.

2. Identify the symmetry groups of other letters of the alphabet.

It is important in the study of symmetry to distinguish between direct and opposite isometries. Briefly, direct isometries are those which are possible rigid movements: so, in the plane, rotations are direct while reflections are opposite, since the latter cannot be performed as rigid movements within the plane. The composite of two opposite isometries is a direct isometry.

It may be proved as a theorem in transformation geometry

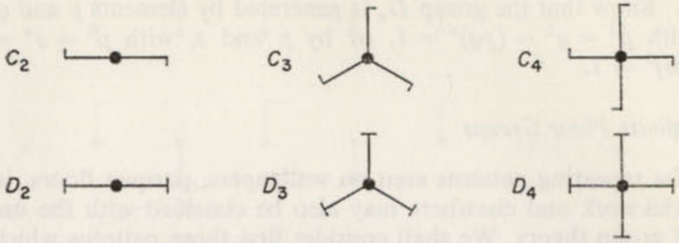


Fig. 30 Symbolic diagrams for some plane symmetry groups.

that *every* isometry of the plane which has a fixed point is either a rotation or a reflection, and this enables us to describe all the possible symmetry groups of finite plane figures. If we count the identity as a rotation through zero angle, all contain rotations: if there are n rotations they will be multiples of $2\pi/n$; others contain, with the rotations, an equal number of reflections, in axes which make angles of π/n with each other. The group of n rotations is cyclic, isomorphic to C_n, and that containing n reflections as well is called the *dihedral group of degree n*, and is denoted by D_n. The diagrams of Figure 30 show some of these.

Exercises

1. In the regular pentagon (Figure 31) show how to transform region 1 (a) into region 5, (b) into region 7, by suitable combinations of the reflections p and q.

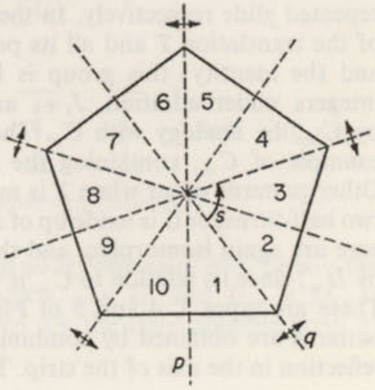


Fig. 31 The regular pentagon and its symmetries.

2. Show that the group D_n is generated by elements p and q, with $p^2 = q^2 = (pq)^n = 1$, or by p and s, with $p^2 = s^n = (ps)^2 = 1$.

Infinite Plane Groups

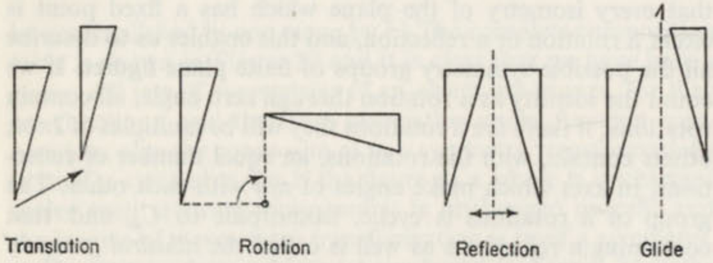


Fig. 32 Isometries of the plane.

The repeating patterns seen on wallpapers, parquet floors, in brickwork and elsewhere may also be classified with the use of group theory. We shall consider first those patterns which repeat along a strip, such as a piece of wallpaper border. In this case geometrical considerations show that the basic movement is either a translation or a glide (see Figure 32). (A glide is the composite of a reflection and a translation in the direction of the mirror). The simplest types of pattern (the first two of Figure 33) have as symmetries just a repeated translation or a repeated glide respectively. In the first case the group consists of the translation T and all its powers, positive and negative, and the identity; this group is isomorphic to the group of integers under addition, $J,+$, and it is sometimes denoted by C_∞, by analogy with C_n. The second case gives another example of C_∞, containing the glide G and all its powers. Other patterns occur when T is made up of two reflections, or two half-turns, or G is made up of one of each. The three groups here are again isomorphic, and this abstract group is denoted by D_∞, since its relation to C_∞ is the same as that of D_n to C_n. These are types 3, 4 and 5 of Figure 33. The remaining two patterns are obtained by combining with the previous ones a reflection in the axis of the strip. This reflection is independent of the translation or glide and so it commutes with these—the groups are the direct products $C_\infty \times D_1$ and $D_\infty \times D_1$. It can be shown that these seven are the only possible ways of repeating a pattern on a strip.

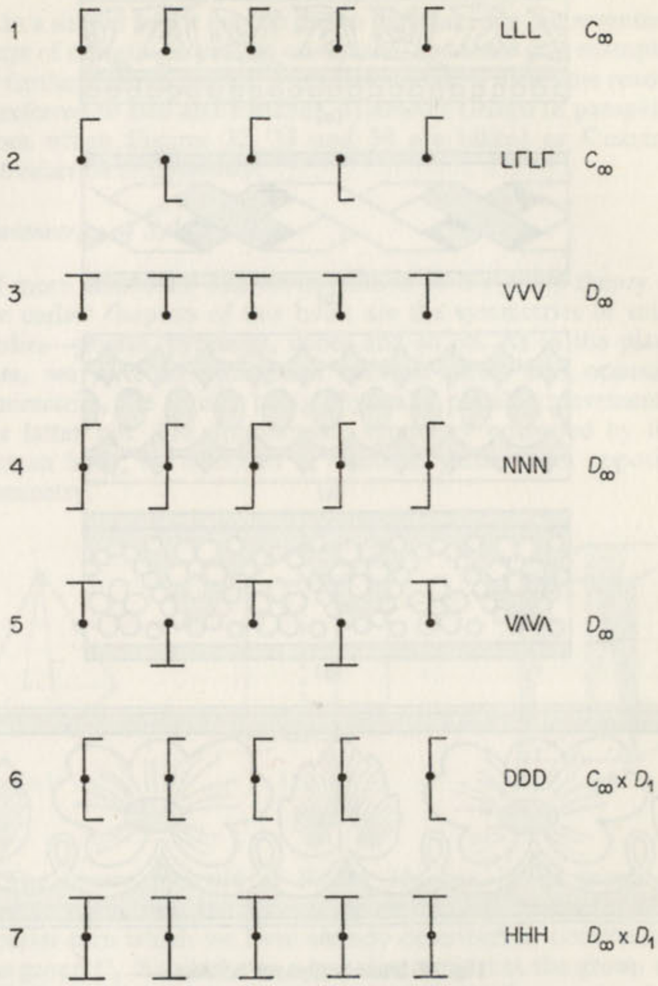


Fig. 33 The seven strip patterns.

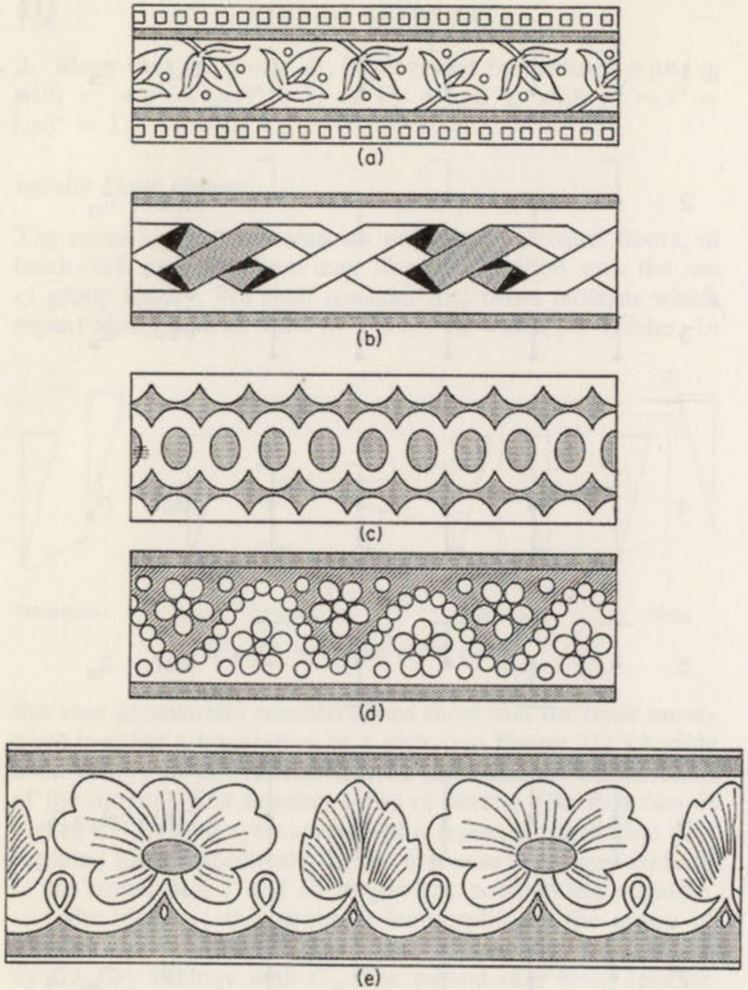


Fig. 34 Some wallpaper borders.

Exercises

1. State generating relations for D_∞ and $D_\infty \times D_1$.

2. Identify the symmetry types of the borders illustrated in Figure 34. (DIfferent answers are possible according to whether or not the shading is taken into account.)

In a similar way it may be shown that there are just seventeen ways of repeating a pattern on a plane, and these give examples of further infinite groups. For a discussion of these, the reader is referred to Bell and Fletcher, *Symmetry Groups* (a pamphlet from which Figures 32, 33 and 34 are taken) or Coxeter, *Introduction to Geometry*.

Symmetries of Solids

Of more immediate interest in relation to the group theory of the earlier chapters of this book are the symmetries of solid bodies—prisms, pyramids, cubes and so on. As in the plane case, we have to distinguish between direct and opposite symmetries, the former being physically possible movements, the latter not. The approximate symmetry possessed by the human body, by reflection in a central plane, is an opposite symmetry.

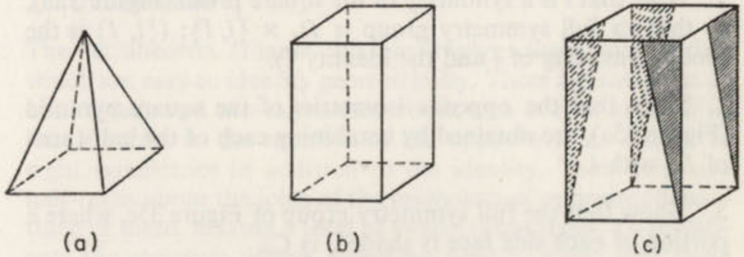


Fig. 35.

The square pyramid of Figure 35a has, as its group of direct symmetries, the four rotations through multiples of a quarter-turn which we have already described as constituting the group C_4. Similarly the n-gonal pyramid has the group C_n as its group of direct symmetries. But what will correspond in three dimensions to the groups D_n in two? We could use this as the symbol for the full group of the pyramid including the reflections in the bisecting planes along with the rotations. But, in practice, D_n is used for the group obtained by adding to the rotations of C_n the n *half-turns* about axes in a plane perpendicular to the axis of the rotations and making angles π/n with each other. This reduces to the familiar two-dimensional D_n, and has the advantage of making the three-dimensional D_n a group consisting entirely of direct isometries. It is, in fact, the group of *direct* isometries of the n-gonal *prism* (Figure 35b shows the case $n = 4$).

It can be shown from geometrical considerations that every isometry of space with a fixed point is either a rotation or a combination of a rotation with the *central inversion* i; this is a transformation which reflects every point of space in the fixed point, so that if the fixed point is the origin of coordinates, the point (x, y, z) becomes $(-x, -y, -z)$. This makes it possible to obtain all symmetry groups which contain opposite isometries as suitable combinations of direct isometry groups with i.

Exercises

1. Note that i is a symmetry of the square prism (Figure 32b), so that its full symmetry group is $D_4 \times \{i, I\}$; ($\{i, I\}$ is the group consisting of i and the identity I).

2. Show that the opposite isometries of the square pyramid (Figure 35a), are obtained by combining each of the half-turns of D_4 with i.

3. Show that the full symmetry group of Figure 35c, where a portion of each side face is shaded, is C_4.

4. Find the symmetry groups which Figure 35c would have with various shadings. (Use a model.)

It is easy to show, on group-theoretical grounds, that if a symmetry group contains both rotations and rotatory inversions (that is, combination of rotations with the central inversion), it must contain an equal number of each, and the set of rotations is a subgroup of the whole group. For if we combine every element with one of the rotatory inversions, r_1 say, the rotatory inversions all become rotations and vice versa. But this set of composites with r_1 must, by the group closure property, be the same set of elements in a different order: thus there must be equal numbers of each type.

SYMMETRIES OF THE TETRAHEDRON AND CUBE

The three-dimensional groups discussed above are all derived from the groups C_n and D_n in two dimensions; apart from these, three more groups, with similar derivatives, exist, and they are associated with the five Platonic regular solids, the tetrahedron, cube, octahedron, dodecahedron and icosahedron.

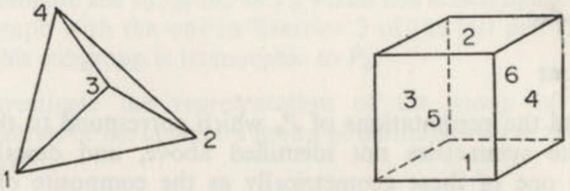


Fig. 36 The regular tetrahedron and the cube.

The Tetrahedron

The tetrahedron (Figure 36) has twelve direct symmetries, which are easy to identify geometrically. There are rotations of a third-turn and two-thirds-turn about each of four axes which join a vertex to the centroid of the opposite face: a total of eight symmetries in addition to the identity. Then there are half-turns about the joins of the midpoints of opposite edges—three of these, making a total of twelve symmetries. To investigate the structure of this group we need to know how these movements combine; this can be discovered from experiments with a marked model, but a simpler method will appear below, as we go on to consider the opposite symmetries. These include reflections in the planes joining each edge to the midpoint of the opposite edge—six of these. The other six opposite symmetries are not so obvious geometrically, but we can discover them by analysing the situation in terms of permutations. Every symmetry of the tetrahedron is equivalent to a permutation of its four vertices. If the vertices are labelled 1, 2, 3, 4, a permutation such as (12) is one of the reflections described above, the third-turn rotations are of the type (234), and the half-turns are of the form (12)(34). It is clear that the direct symmetries are even permutations (see p. 102) and that the reflections such as (12) are odd. Moreover, *every* transposition like (12) is a symmetry of the tetrahedron, and so therefore is every composite of two or more of these. And since every permutation may be expressed as a composite of transpositions (p. 102), every permutation of $\{1, 2, 3, 4\}$ is a symmetry of the tetrahedron, the direct symmetries corresponding to the even permutations, the opposite symmetries to the odd. Thus the full symmetry group of the tetrahedron is isomorphic to the group P_4, and the group of direct symmetries to the alternating group A_4.

Exercises

1. Find the permutations of P_4 which correspond to the six opposite symmetries not identified above, and describe a typical one of these geometrically as the composite of two simpler geometrical transformations.

2. Express the symmetries of the tetrahedron as permutations of the four *faces*, and show which of these corresponds to which permutation of the vertices.

3. Investigate the representation of the symmetry groups of the tetrahedron as permutations of the six edges.

The Cube

As in the case of the tetrahedron, the symmetries of the cube can be investigated by considering them as permutations of its vertices, or faces, or edges. The simplest of these ways is to consider faces, as there are only six of them, compared with eight vertices and twelve edges. If the cube is labelled as in Figure 36, using the pairs 12, 34, 56 for pairs of opposite faces, we can note some of the direct symmetries as follows:

Rotations through multiples of a quarter-turn about axes joining centres of opposite faces—(3546), (34)(56), (3645) and two similar sets—nine symmetries;

Half-turns about axes joining midpoints of opposite edges—(36)(45) and five similar to this—total, six symmetries;

Rotations through multiples of a third-turn about diagonals of the cube—(135)(246), (153)(264) and three similar sets—total, eight symmetries.

This gives, with the identity, twenty-four direct symmetries; this is the correct total, since a cube has twenty-four different positions: any one of the six faces can be uppermost, and then any one of the four vertical faces may face the front.

Exercises

1. Compare the subgroup of P_6 which has arisen in the above paragraph with the one in Exercise 3 of the last set. Deduce that this subgroup is isomorphic to P_4.

2. Investigate the representation of the group of direct symmetries of the cube as a group of permutations of the eight vertices.

Exercise 1 suggests that it may be possible to represent the direct symmetries of the cube by a simpler permutation group using only four symbols. This is indeed the case, if the four elements are the four diagonals of the cube. It is clear that every symmetry transformation of the cube must correspond to some permutation of the diagonals, but we need to verify that each permutation of the diagonals corresponds to one and only one direct symmetry. In other words, we must consider whether it is possible to make a symmetry transformation which leaves all four diagonals in their original positions. (We do not say 'fixed', for a diagonal could be reversed in direction but still occupy its original position, in the sense that diagonal 1 would still be where diagonal 1 was before.) It may be seen that this is not possible, for a rotation which kept a diagonal fixed would have to be about that diagonal, and one which reversed a diagonal would have to be about an axis at right angles to the diagonal. Thus a rotation which left all four diagonals either fixed or reversed would have to be about an axis which was either along or at right angles to each one of them, and it is immediately clear from the geometry of the figure that this is impossible. The direct symmetry group of the cube is therefore isomorphic to P_4.

It follows that the full group including opposite symmetries cannot be expressed in terms of permutations of fewer than six elements, and this does not give a very useful characterization of the group. Since, however, every opposite symmetry is the composite of a direct symmetry with the central inversion *i*, and *i* commutes with each such symmetry, we may characterize the full group as $P_4 \times \{i, I\}$, which is abstractly equivalent to $P_4 \times C_2$. This group is sometimes denoted by $\overline{P}_4$.

Exercises

1. Show that the central inversion of the cube (12)(34)(56) commutes with each direct symmetry.

2. Note that in this case even and odd permutations do not correspond to direct and opposite symmetries respectively.

3. Pursue the relationship indicated in Exercise 1 of the last set by showing how a regular tetrahedron may be placed inside a cube so that each edge of the tetrahedron is a diagonal of a face of the cube.

Other Regular Solids

Figure 37 shows how an octahedron may be placed in relation to a cube so that each vertex of the former is the centre of a face of the latter, and it can be seen from the figure that each face of the octahedron corresponds to a vertex of the cube. Thus the symmetry group of the cube, represented as a group of permutations of the faces, is also the symmetry group of the octahedron, represented as a group of permutations of the vertices. The two solids have exactly the same symmetry. A similar relationship exists between the remaining two regular

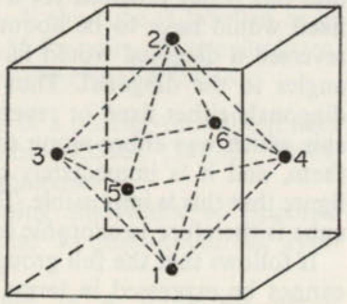


Fig. 37 The cube and octahedron.

solids, the dodecahedron and the icosahedron. The former has twelve faces and twenty vertices, the latter twenty faces and twelve vertices, and both have thirty edges. Examination of models of these will make it clear that they correspond to each other in the same way as the cube and the octahedron, and that they have the same symmetry groups. It is also a fairly easy matter to find how many symmetries they have; thinking of the dodecahedron, it may be placed on any one of the twelve faces, and may then be rotated into any one of five positions: total, sixty positions. This is the number of direct symmetries, and the number of opposite symmetries is therefore a further sixty. It is a plausible guess that the groups we have here might be isomorphic to P_5 and A_5, and this can indeed be shown to be the case. We use the icosahedron for this demonstration.

The twenty faces of this figure can be labelled with the numbers 1, 2, 3, 4, 5 in such a way that the five triangles meeting at any vertex all have different numbers (see Figure 38, where the solid is opened out and distorted so that all faces can be seen except the one opposite the face in the centre. It is convenient to have a model whose faces are coloured with five different colours in this way.) The four faces labelled 1 are then distributed symmetrically over the icosahedron, lying in the faces of an (imaginary) regular tetrahedron. The faces

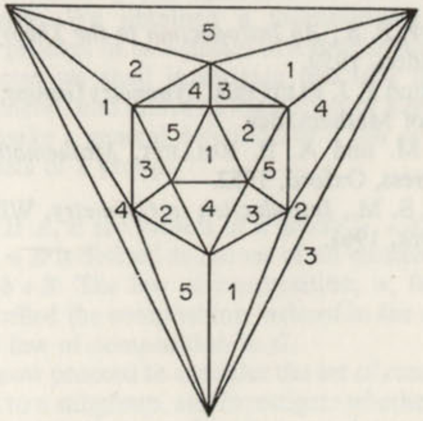


Fig. 38 A possible numbering of the faces of the icosahedron.

opposite to these four will be labelled 2, 3, 4, 5 and will lie in the faces of another regular tetrahedron. The symmetries of these tetrahedra are symmetries of the icosahedron; we have shown above that the symmetry groups of the tetrahedron are the group of all permutations of its four faces, for all symmetries, and the subgroup of even permutations for the direct symmetries. When these symmetry transformations are applied to the tetrahedra identified in the icosahedron, they are represented by the permutations of $\{2, 3, 4, 5\}$, with 1 remaining unchanged. Since what we have said about the faces labelled 1 is equally true of those labelled with any other number, we have in fact considered the $5 \times 24 = 120$ permutations of P_5, and have shown that each of them corresponds to a different symmetry of the icosahedron, and the even permutations to the direct symmetries. As 120 is the total number of symmetries, we can identify the full and direct symmetry groups with P_5 and A_5 respectively.

Exercise

Identify the symmetry groups of some of the semi-regular solids.

REFERENCES

ALEXANDROFF, P. S., *An Introduction to the Theory of Groups*, Blackie, London, 1959.

BELL, A. W. and T. J. FLETCHER, *Symmetry Groups*, Association of Teachers of Mathematics.

CUNDY, H. M. and A. P. ROLLETT, *Mathematical Models*, Clarendon Press, Oxford, 1952.

COXETER, H. S. M., *Introduction to Geometry*, Wiley, London and New York, 1961.

9

QUOTIENT GROUPS AND FIELD EXTENSIONS

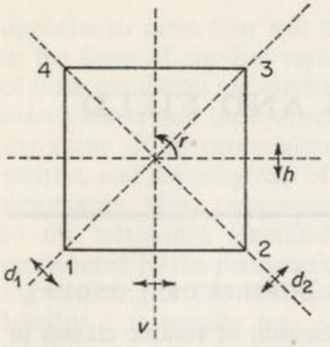
LAWS OF COMPOSITION FOR SUBSETS OF A GROUP

We have already met, in our discussion of residue classes in Chapter 5, one example in which the subsets of a group formed a group among themselves, with a law of composition derived from that of the original group in an obvious way. We considered classes **0**, **1**, **2**, **3**, **4**, **5** consisting respectively of integers which left remainders 0, 1, 2, 3, 4, 5 after division by 6, and the *sum* of two classes was defined as that class containing all the sums of pairs of elements, one taken from each of the classes being combined: in symbols, $\mathbf{A} + \mathbf{B}$ was the class $\mathbf{C}$ containing $a + b$, for all $a \varepsilon \mathbf{A}$ and $b \varepsilon \mathbf{B}$. The product of two classes was defined in a similar way. To ensure that such a definition was valid, we had to verify that all $a + b$, with $a \varepsilon \mathbf{A}$ and $b \varepsilon \mathbf{B}$, were in fact in the same one of the five classes being considered. We thus obtained a group $J_6, +$, whose elements were the residue classes. We obtained a multiplicative group too, defining the product of two classes in a similar way to the sum. In this chapter we shall investigate this idea in relation to groups in general, and arrive at the notion of a *quotient group*.

We first make a general definition of a law of composition for the subsets of a group.

Definition If A, B are subsets of a group $G, *$, then the composite set $A * B$ is defined as the set of all elements $a * b$, with $a \varepsilon A$ and $b \varepsilon B$. The law of composition, $*$, for the sets is sometimes called the composition *induced* in the set of subsets of G by the law of composition in G.

We shall now proceed to consider the set of *cosets* of a group with respect to a subgroup, and investigate whether the induced law of composition among the cosets produces a structure of



$I = (1)$	$h = (14)(23)$
$r = (1234)$	$v = (12)(34)$
$r^2 = (13)(24)$	$d_1 = (24)$
$r^3 = (1432)$	$d_2 = (13)$

		S				C			
		I	r	r^2	r^3	h	v	d_1	d_2
S	I	I	r	r^2	r^3	h	v	d_1	d_2
	r	r	r^2	r^3	I	d_1	d_2	v	h
	r^2	r^2	r^3	I	r	v	h	d_2	d_1
	r^3	r^3	I	r	r^2	d_2	d_1	h	v
C	h	h	d_2	v	d_1	I	r^2	r^3	r
	v	v	d_1	h	d_2	r^2	I	r	r^3
	d_1	d_1	h	d_2	v	r	r^3	I	r^2
	d_2	d_2	v	d_1	h	r^3	r	r^2	I

	S	C
S	S	C
C	C	S

Fig. 39 The group D_4.

any interest. We consider first the group D_4, the group of symmetries of the square. We shall need the full group composition table for reference, and this is shown in Figure 39. The reader is recommended to work this table out for himself, as an exercise; this may be done by experimenting with a cut-out square, but it is probably more convenient to express each of the eight symmetries as a permutation, and to calculate the composites from these. If the cut-out square is used, care must be taken to interpret h,v and so on consistently, either with respect to axes fixed in space or with respect to axes fixed in the square, remembering in the second case that r is an anticlockwise rotation when seen from the face of the square which was originally upward, so that when the square is turned over, r is anticlockwise as seen from below. In the table given, the axes are taken to be fixed in space, the numbers are allotted to the four space-positions and (123) means that the vertex which was in position 1 moves to position 2, and vh (v row, h column) means h first, then v. Other conventions may give a different table.

To form cosets, let us choose first the subgroup $S = \{1, r, r^2, r^3\}$, There is only one other coset, and it consists of the remaining elements of the group, $C = \{h, v, d_1, d_2\}$. The composite set SC will consist of the sixteen elements obtained by combining an element of S with one of C:

$$Ih, Iv, Id_1, Id_2;\ rh\ rv, rd_1, rd_2;\ r^2h, \ldots$$

and so on. These are, in turn

$$h, v, d_1, d_2;\ d_1, d_2, v, h, \ldots$$

and so on. All these elements prove to be in the coset C, so we may say that $SC = C$. It may be similarly verified that $C^2 = S$, and that $CS = C$, and $S^2 = S$—this last follows from the fact that S is a subgroup. Thus S and C form a two-element *group*, whose table is shown in Figure 39; the figure also indicates how the composition table for S and C appears in the original group table, in the way in which it divides into four blocks. When a group G is divided by a subgroup S into cosets, and the cosets form a group under the induced law of composition, the resulting group is called the *quotient group* of G by S, written G/S.

In the above example, the quotient group had only two elements. Can we derive a larger quotient group from D_4, by using a smaller subgroup? We shall try the subgroup $S = \{I, h\}$: this should give a four-element quotient group. The left-cosets are $\{r, d_1\}$, $\{r^2, v\}$, $\{r^3, d_2\}$: call them P, Q, R respectively. We consider now the composites, taking PR first. This gives the set $\{r^4, rd_2, d_1r^3, d_1d_2\}$ which is $\{I, h, v, r^2\}$—which is not one of the cosets! So we have no quotient group.

Why has the process broken down? In the previous example, the sixteen composites reduced to four; why do not these four reduce to two?

To see why, we must write the cosets so as to show how they were obtained: they may be written as S (the subgroup), d_1S, vS, and d_2S (or we could have written them S, rS, r^2S, r^3S). The composite set $d_1S \cdot d_2S$ is then the set of all elements of the form $d_1s_1d_2s_2$, where s_1, s_2 run through all elements of S. If these are all to reduce to the elements of one coset, it will be the coset of d_1d_2 since $s_1 = I$, $s_2 = I$ is one pair of values to be included; so $d_1s_1d_2s_2$ must be of the form $d_1d_2s_3$, where s_3 is some member of S. By cancelling d_1 and multiplying on the right by $\bar{s}_2$, we have

$$\begin{aligned} d_1s_1d_2s_2 &= d_1d_2s_3 \\ \Rightarrow \quad s_1d_2 &= d_2s_3\bar{s}_2 \\ &= d_2s_4 \end{aligned}$$

which if it is to be true for all s_1, is equivalent to saying that the sets d_2S and Sd_2 must be the same, in other words the right and left-cosets of d_2 with respect to S must be the same. This condition is clearly fulfilled in the previous example, since as the subgroup contains half the elements of the group, there is only one possible other coset. A subgroup, with respect to which the right- and left-cosets of each element are the same, is called an invariant subgroup, or a normal subgroup; we shall prove below that the cosets with respect to a normal subgroup always form a quotient group.

Definition A *normal subgroup* N of a group G is a subgroup such that the left-coset of each element of G is the same as the right-coset of the same element: in symbols,

$$\forall\, a \,\varepsilon\, G, \quad aN = Na.$$

(Note that this states that aN and Na are the *same set*; there will in general be individual elements n of N such that $an \neq na$.)

Theorem If N is a normal subgroup of a group G, the set of cosets of G with respect to N forms a group, the law of composition of the cosets being the law induced by the law of the group G. This group is called the *quotient group* of G by N, written G/N.

Proof

(i) *Closure*: the composite $(aN)(bN)$ of two cosets reduces as follows:

$$\begin{aligned} (aN)(bN) &= aNbN \\ &= abNN \\ &= (ab)N \end{aligned}$$

which is the coset of ab.

(ii) *Associativity* follows from the associativity of the elements of G, since $(an_1)(bn_2)(cn_3)$ is a composite of elements of G.

(iii) The *identity* element is the subgroup N, since

$$(aN)N = aN = N\,(aN).$$

(iv) The *inverse* of aN is the coset $\bar{a}N$, since

$$\begin{aligned} (aN)(\bar{a}N) &= a\bar{a}N \\ &= N. \end{aligned}$$

The four group axioms are thus satisfied.

Exercises

1. Divide the group D_4 by the subgroup $\{I, r^2\}$, and find to which of the two abstract groups of four elements the quotient group is isomorphic.

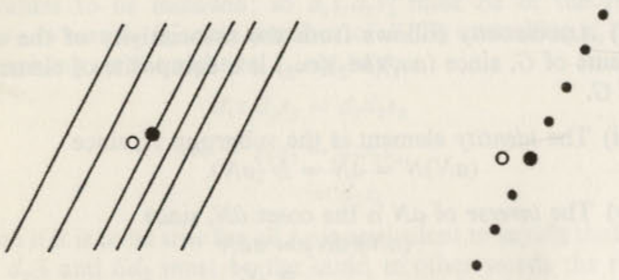
2. Find a normal subgroup of D_3 and form the quotient group.

3. Show that the quotient group of the group of integers under addition by the subgroup of multiples of 5 is the group $J_5, +$.

4. Divide P_4 by a normal subgroup isomorphic to P_2 and identify the quotient group.

Further examples of Quotient Groups

The points of a plane may be made to become the elements of a group if we label them by coordinates (x, y), $(x,y \,\varepsilon\, R)$, and define a composition $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$. This is the law usually called vector addition, and we shall write $\boldsymbol{x}$ for the pair (x, y) as is customary. If O is the origin of coordinates, and $\boldsymbol{a}$ is the pair representing a point A, then the set $\{\lambda \boldsymbol{a}\}$, for all $\lambda \varepsilon R$, is a subgroup of the whole group of pairs P, and so is the set $\{k\boldsymbol{a}\}$, where $k \varepsilon J$. These are illustrated in Figure 40, which also shows some *cosets of the subgroup* $\{\lambda \boldsymbol{a}\}$, which are the lines parallel to the line $\{\lambda \boldsymbol{a}\}$. These cosets form a quotient group (the group P is commutative), in which



Some cosets $\boldsymbol{b} + \{\lambda \boldsymbol{a}\}, \lambda \varepsilon R$

The subgroup $\{k\, \boldsymbol{a}\}, k\ \varepsilon\ J$

Fig. 40 Cosets in the group, P, of points of the plane.

the 'sum' of two lines is the line containing any point obtained by adding one point of each of the two component lines.

Another example is provided by the group of functions $x \to ax + b$: we leave the details to the reader. (See the exercises below.)

Exercises

1. Find the cosets into which the group P of points of the plane is divided by the other subgroups mentioned above.

2. The set of functions $f_{ab}: R \to R: x \to ax + b$, for all $a, b \varepsilon R$, forms a group both under function composition and under addition (see p. 66). Show that the subset $f_{a0}: x \to ax$, for all a, forms a subgroup in each of these cases, and that the subgroup is normal in one of them. In this case, find the quotient group.

3. Show that the quotient group of R, + by J, + is a group which might be called the 'additive group of real residues modulo 1'.

4. Consider the group of permutations of $\{a, b, c, d\}$ which leave the cross-ratio $(a-c)(b-d)/(a-d)(b-c) = \lambda$ invariant. Show that it is a normal subgroup of P_4; find the quotient group, and show the isomorphism between this and the group of values of the cross-ratio, λ, $1/\lambda$, $1-\lambda$, $1/(1-\lambda)$, $(\lambda-1)/\lambda$, $\lambda/(\lambda-1)$. (Hint: What is the law of composition?)

FIELD EXTENSIONS

We include here a short section on the extension of fields, since it has relevance to several topics discussed in this book. It may be regarded in some ways as the reverse process to that of forming a quotient structure—it will be seen that when a field has been extended the original field may be recovered as a quotient. We take up the discussion of fields at the point at which we left it in Chapter 5, having proved that a quadratic equation with coefficients in any field F may be reduced to the form $(x - h)^2 = p$, where p is an element of F. Whether or not the solution of the equation can be completed from this point depends on whether p has a square root in the field F. If F is the real field R, and p is negative, there is no value for x in F, and this leads to the extension of R to form the complex field C. Instead of pursuing this example, we shall consider the case when F is the finite field J_5. In this case $p = 0$, 1 or 4 gives a solution for x, but there is no solution if $x = 2$ or 3. We may therefore consider an element t, such that $t^2 = 2$: we note that this provides a square root for 3 as well, since $(2t)^2 = 4.2 = 3$. We can prove further that if t is 'adjoined' to the field J_5, that is to say, if we include all sums and products obtained from t and the elements of J_5, then the resulting structure is also a field. The set consists of

0	1	2	3	4
t	$1 + t$	$2 + t$	$3 + t$	$4 + t$
$2t$	$1 + 2t$	$2 + 2t$	$3 + 2t$	$4 + 2t$
$3t$	$1 + 3t$	$2 + 3t$	$3 + 3t$	$4 + 3t$
$4t$	$1 + 4t$	$2 + 4t$	$3 + 4t$	$4 + 4t$

There are 25 elements in all, and we may express them as $\{a + bt\}$, with $a,b \varepsilon J_5$. It is clear that the sum, product and difference of any two of these is of the same form, and that the compositions are associative, commutative and distributive, $+$ over $\cdot$; the only condition for a field that requires attention is the existence of multiplicative inverses for all elements except 0. But we have

$$\frac{1}{a + bt} = \frac{a - bt}{(a + bt)(a - bt)} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}t,$$

which is of the required form. $a^2 - 2b^2$ cannot be 0 since $a^2 = 2b^2$ would imply $(a/b)^2 = 2$, which is impossible. This completes the proof.

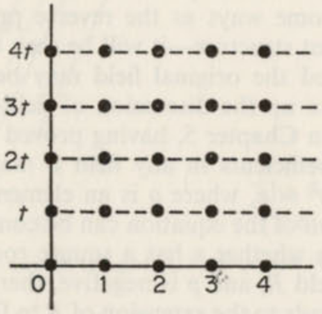


Fig. 41 The field $J_5(t)$, an additive subgroup $\{a\}$, $a \varepsilon J_5$, and cosets.

The additive subgroups of this field include $\{a\}$, with $a \varepsilon J_5$, and $\{bt\}$, with $b \varepsilon J_5$, and the reader may easily verify that the quotient group in each of these cases is isomorphic to J_5; if the field elements are represented by the points of Figure 41, the cosets are represented by the lines, with addition defined in the obvious way. When we consider the multiplicative group contained in the field, we have now 24 elements, 0 being excluded, and the quotient groups are more interesting. The subgroup $S = \{1, 2, 3\ 4\}$, for example, divides the group into six cosets, which may be written S, tS, $(1 + t)S$, $(2 + t)S$, $(3 + t)S$, $(4 + t)\,S$. These form a quotient group which is cyclic (the only other abstract group of order 6 is non-commutative), and it may be verified that $(3 + t)S$ is a generator of this group.

Exercises

1. Verify the statements of the above paragraph.

2. Find the structure of the 24-element multiplicative group mentioned above.

3. Find a subgroup of 3 elements of this multiplicative group, and predict the structure of the resulting quotient group.

4. Make a similar investigation starting with the field J_3.

Further Field Extensions and Vector Spaces

All cubic equations in a field F can be reduced to the form $x^3 + ax + b = 0$, with a and b in F, but, as in the case of quadratics, the complete solution of the equation depends on the existence of a root of this reduced equation in the field F. If F is J_5, all such equations with $a = 0$ or $b = 0$ are soluble, but, for example, $x^3 + x + 1 = 0$ is not. J_5 may be extended by adjoining to it a root of this equation, r say; this generates a set consisting of all elements of the form $a + br + cr^2$, with a, b, c in J_5, 5^3 elements in all. (Any elements which contained r^3 could be reduced by using the relation $r^3 + r + 1 = 0$ to elements of the form $a + br + cr^2$.) As before, it may be proved that this extended set $J_5(r)$ is a field. Subgroups and quotient groups of this field may be investigated as above. Readers acquainted with vector spaces will observe that the fields of these last two paragraphs are vector spaces over J_5, of dimension 2 and 3 respectively.

Exercises

1. Show that $J_5(r)$ is a field.

2. Identify some of the subgroups and quotient groups in the field $J_5(r)$.

SYMMETRIES OF FIELDS—AUTOMORPHISMS

Automorphisms of groups were discussed in Chapter 7; a field automorphism is a one-one mapping of a field onto itself such that both field compositions are preserved. As we observed before, the set of automorphisms, which is itself a group, gives a measure of the symmetry of the system. The transformation $t \to 4t$ is an automorphism of the field J_5: it is a one-one mapping of the field onto itself, it preserves addition, and also multiplication—any product not involving t^2 is clearly preserved, and since $t^2 = (4t)^2 = 2$, so is any product which does involve t^2.

The group of automorphisms in this case consists of the identity mapping and $t \to 4t$, the square of which is the identity.

Exercises

1. Show that the complex field, regarded as an extension R(i) of the real field, has the automorphism i → −i.

2. Show that the field $J_5(r)$ defined above has the one non-identical automorphism $r \to r^2$; and that this gives an automorphism group isomorphic to C_2.

3. Extend some of the ideas of the last two sections to other finite fields, such as J_3, J_2.

Field extensions and their automorphisms play an important part in the Galois theory of the solubility of equations, including the proof of the insolubility, by radicals, of the general equation of the fifth degree, and we have already mentioned that it was in this field that the theory of groups was first used. We cannot treat this subject fully within the scope of this book. We have, however, tried to show something of the inter-relationship of the commoner algebraic structures, and particularly to show the breadth of application of the idea of the group.

REFERENCES

BIRKHOFF, G. and S. MACLANE, *A Survey of Modern Algebra*, second edition, Macmillan, New York, 1953.

SAWYER, W. W., *A Concrete Approach to Abstract Algebra*, W. H. Freeman, San Francisco, 1959.

ANSWERS TO EXERCISES

CHAPTER 6

Page 87.

1. $\{1, 2, 4\}$, $\{3, 5, 6\}$.

2. $\{1, 10\}$, $\{2, 9\}$, $\{3, 8\}$, $\{4, 7\}$, $\{5, 6\}$.

3. $\{0, 3\}$, $\{1, 4\}$, $\{2, 5\}$.
$\{0, 2, 4\}$, $\{1, 3, 5\}$.

4. $\{1, p\}$, $\{w, q\}$, $\{w^2, v\}$.

Page 94.

3. J_{20}^8 and J_{30}^8 are both isomorphic to $C_4 \times C_2$, being generated by 3 and 11, and 7 and 11, respectively; J_{24}^8 is isomorphic to $C_2 \times C_2 \times C_2$, having generators 5, 7, 13.

CHAPTER 7

Page 97.

1. (13), (13), (132).

Page 98.

1. (123).

2. (123), (132), (132), (123).

3. (1), (1), (132), (1).

5. See p. 110; $Q3$ and $Q4$ must be classified together.

6. (21)(24)(23), (32)(31)(34), (43)(42)(41).

7. (1432), (13)(24).

8. C_4.

Page 100.

4. Subgroups generated by (1243), (1324).

6. Generated by (1243) and (14), (1324) and (12).
No; (1), (12)(34), (13)(24), (14)(23) appear in all three, while permutations of order 3 do not appear at all.

9. Generated by (123 . . . n).

Page 102. (*In most cases the answers given are not the only correct ones.*)

1. (13)(14)(12).

2. (1342), (12)(14)(13).

3. (12)(14)(13)(25)(24)(21), (15)(14)(13)(12), (24)(23)(15)(14).

Page 105.

1. (1), (123), (132);
The permutations of classes $Q0$, $Q2$, $Q3$ on p. 110.

3. (132)(134), (123)(124), (124)(123).

4. No; it is an odd permutation and 3-cycles are even.

Page 107.

1. $a = (1a)(bc)$, $b = (1b)(ac)$, $c = (1c)(ab)$.

2. $r = (1rr^2r^3)$, $r^2 = (1r^2)(rr^3)$, $r^3 = (1r^3r^2r)$.

Page 108.

2. D_4.

3. Inner automorphisms A_4, full group P_4.

Page 110.

1. Transformed by ↓	1	w	w^2	v	p	q	Inner Automorphism
w	1	w	w^2	p	q	v	(vpq)
w^2	1	w	w^2	q	v	p	(vqp)
v	1	w^2	w	v	q	p	$(ww^2)(pq)$
p	1	w^2	w	q	p	v	$(ww^2)(vq)$
q	1	w^2	w	p	v	q	$(ww^2)(pv)$

Group D_2

3. Transformed by ↓	1	r	r^2	r^3	h	v	d_1	d_2	Inner Automorphism
r	1	r	r^2	r^3	v	h	d_2	d_1	$(hv)(d_1d_2)$
r^2	1	r	r^2	r^3	h	v	d_1	d_2	(1)
r^3	1	r	r^2	r^3	v	h	d_2	d_1	$(hv)(d_1d_2)$
h	1	r^3	r^2	r	h	v	d_2	d_1	$(rr^3)(d_1d_2)$
v	1	r^3	r^2	r	h	v	d_2	d_1	$(rr^3)(d_1d_2)$
d_1	1	r^3	r^2	r	v	h	d_1	d_2	$(rr^3)(hv)$
d_2	1	r^3	r^2	r	v	h	d_1	d_2	$(rr^3)(hv)$

Group D_2

6. A_4: take $Q0$, $Q2$, $Q3$.

C_4: (1234), (13)(24), (1432).

D_2: $Q3$ or (12), (34), (12)(34); other sets possible.

D_3: All permutations not containing 4.

Page 111.

1, 2. p. 108, Ex. 1: Inner automorphisms (1) only (cf p. 110 Ex. 2) Ex. 2: D_3: All automorphisms inner. D_4: Inner D_2. Ex. 3: Inner A_4.

CHAPTER 8

Page 114.

1. A, B: D_1; N, parallelogram: C_2.

Page 115.

1. (a) $(pq)^3$; (b) $(pq)^2$.

Page 118.

1. $p^2 = q^2 = 1, (pq)^n \neq 1$ for any n; same with $r^2 = 1, pr = rp, qr = rq$.

2. (a) Type 2 (b) 1 or 4 (c) 7 (d) 3 or 5 (e) 3.

Page 122.

1. The six 4-cycles of P_4 (see p. 110); (1234) = (12)(234), a third-turn about the axis through vertex 1 followed by reflection in the perpendicular bisector plane of edge 12.
2. If vertex A is opposite face a, and so on, (AB) is the same transformation as (ab).

CHAPTER 9

Page 131.

1. D_2.

2. Isomorphic to C_3, C_2.

4. Isomorphic to P_3.

Page 132.

1. $\{b+ka\}$, for fixed b and all $k \varepsilon J$, is a coset.

2. Under addition: quotient group is isomorphic to R, $+$.

4. Function composition; λ is the identity.

Page 135.

2. Cyclic; $3+t$ is a generator.

3. $\{1, 2+2t, 2+3t\}$. Quotient group isomorphic to C_8.

BIBLIOGRAPHY

For other approaches to number systems see Moakes, *The Core of Mathematics*, which treats the subject more briefly than we do here; Thurston, in *The Number System*, gives a more extensive and more rigorous study, and Birkhoff and Maclane adopt a different point of view, starting by stating axioms for an integral domain and adding further axioms one by one until the resulting system has all the properties of the set of integers.

Sawyer's *A Concrete Approach to Abstract Algebra* provides an alternative treatment of finite arithmetics, of their extension to complex finite fields, and includes also work on extended fields regarded as vector spaces, leading to a proof of the impossibility of trisecting the angle. This would form a suitable continuation of the work of Chapter 9 of this book.

For further reading on group theory, Grossman and Magnus, *Groups and their Graphs* is an easy book which includes most of the group theory treated in this book and has chapters on one or two different applications of the theory; Ledermann's *The Theory of Finite Groups* is more advanced but well supplied with examples. Papy's *Groups* is of a similar standard and has a great many good examples but the terminology is less familiar and the notation more abstract.

The symmetries of plane figures are dealt with in Jeger's *Transformation Geometry* (No. 1 of this series), but for solids one must consult Coxeter's *Introduction to Geometry*, a masterly survey of the whole subject which includes brief but excellent sections on the symmetry topics of this book. Bell and Fletcher, *Symmetry Groups*, is a short pamphlet summarizing the symmetries of plane and solid figures, including patterns which repeat along a strip and over a plane.

Some Lessons in Mathematics includes classroom material to introduce finite arithmetics, symmetry and groups, and a section on transformation geometry.

MOAKES, A. J., *The Core of Mathematics*, Macmillan, London, 1964.

THURSTON, H. A., *The Number System*, Blackie, London, 1956.

BIRKHOFF, G., and S. MACLANE, *A Survey of Modern Algebra*, second edition, McGraw Hill, New York, 1953.

SAWYER, W. W., *A Concrete Approach to Abstract Algebra*, W. H. Freeman, London, 1959.

GROSSMAN, I., and W. MAGNUS, *Groups and their Graphs*, L. W. Singer and Co., New York, 1964.

LEDERMANN, W., *The Theory of Finite Groups*, fifth edition, Oliver and Boyd, 1964.

PAPY, G., *Groups*, Macmillan, London, 1964.

JEGER, M., *Transformation Geometry* (uniform with this volume), Allen & Unwin, London, 1966.

COXETER, H. S. M., *An Introduction to Geometry*, John Wiley, New York, 1961.

BELL, A W., and T. J. FLETCHER, *Symmetry Groups*, Association of Teachers of Mathematics, Nelson, Lancs., 1964.

FLETCHER, T. J. (Editor), *Some Lessons in Mathematics*, Cambridge University Press, London, 1964.

INDEX

MATHEMATICAL STUDIES: 1

TRANSFORMATION GEOMETRY

MAX JEGER

Translated by A. Deicke and A. G. Howson

In 1872 Felix Klein, speaking at the University of Erlangen, suggested that various geometries should be distinguished by the groups of transformations under which their propositions remain valid. This exciting and important idea has had many repercussions in the world of mathematics, and recently its effects have been felt in the school classroom, an outstanding feature of new mathematics syllabuses being the inclusion of an approach to geometry based on the study of such plane transformations as rotations and reflections. This study is doubly profitable, for not only do transformations help to throw geometric properties into sharp relief, but they also provide a fascinating introduction to group theory. Both of these aspects are given due consideration in Mr. Jeger's book, which was described in *Mathematics Teaching* as 'perhaps the best development of school geometry from the group point of view which is to be found anywhere.'

Readers new to this type of geometry will be surprised too by the power and versatility of transformations and by the way that they can be used to solve many different types of construction problems in addition to such well-known results as the nine-point circle. The second aspect is of equal interest. The ways in which transformations are related and the groups they form are investigated. It is shown how the reflections generate all the other congruence-preserving transformations, and that, if in addition enlargements are considered, the result is a new group—the group of similarities which characterises Euclidean geometry. Finally, a look is taken at the geometry associated with affine transformations.

This book is very suitable for sixth formers and undergraduates who want a not too abstract approach to groups, for teachers who want a compact and workable account of this kind of geometry, and for students in Colleges of Education as an introduction to new ideas in algebra and geometry.

Cloth 25s. net
Paper 15s. net

THE WORLD OF MATHEMATICS

JAMES R. NEWMAN (*Editor*)

4 Volumes

Cloth 7 gns. net the set
Paper 21s. net each volume

books that matter

A. W. BELL

ALGEBRAIC STRUCTURES

GEORGE
ALLEN
AND
UNWIN